The Football Gene Project: A Multi-Disciplinary Investigation into the Association of Genetic Polymorphisms with Phenotypes in Football-Specific Contexts

Alexander B.T. McAuley

A thesis submitted in partial fulfilment of the requirements of Birmingham City University for the degree of Doctor of Philosophy

Faculty of Health, Education and Life Sciences

June 2023

Abstract

The focus of current genetic research in sport is on further understanding genotype-phenotype relationships. However, the majority of genetic research has centred on individual sports, which means there is currently a lack of studies on team sports such as football. Since sports vary in their contextual demands, it is likely there are also significant distinctions at the molecular level between athletes of different sports. As such, genetic associations with other sports cannot be generalised to a football-specific context and need to be investigated independently. Accordingly, the overarching aim of this thesis was to investigate the association of genetic polymorphisms with football phenotypes.

The second section of this thesis synthesised existing genetic association research in football. This facilitated the identification of specific methodological limitations and gaps in this research field that could be addressed in the subsequent experimental chapters. There was little evaluation regarding the extent of genetic testing in football, as well as a lack of cross-sectional studies with youth cohorts investigating quantitative traits and development. Following an independent meta-analysis associations were shown between *ACTN3* (rs1815739), *ACE* I/D, and athlete status. An additional narrative synthesis also found *ACTN3* (rs1815739), *ACAN* (rs1516797), and *VEGFA* (rs2010963) may be associated with injury susceptibility.

The first experimental study assessed the prevalence of genetic testing in professional football, with only 10% of coaches, practitioners, and players reporting they had utilised genetic testing. The second, third, fourth, and fifth experimental studies examined the association of several polymorphisms, with technical, psychological, physiological, and age phase phenotypes, respectively, in English academy football players. Significant associations were found between individual polymorphisms, polygenic profiles, and football phenotypes in

all studies. As such, the results of this thesis suggest inter-individual genetic variation does influence football phenotypes and has identified several novel associations that warrant further investigation in larger independent football cohorts.

Publications

McAuley, A. B. T., Baker, J., & Kelly, A. L. (2021). How nature and nurture conspire to influence athletic success. In A. L. Kelly, J. Côté, M. Jeffreys, & J. Turnnidge (Eds.), *Birth advantages and relative age effects in sport: Exploring organizational structures and creating appropriate settings* (pp. 159–183). Routledge.

McAuley, A. B. T., Baker, J., & Kelly, A. L. (2022). Defining "elite" status in sport: From chaos to clarity. *German Journal of Exercise and Sport Research*, *52*(1), 193–197. https://doi.org/10.1007/s12662-021-00737-3

McAuley, A. B. T., Hughes, D. C., Tsaprouni, L. G., Varley, I., Suraci, B., Baker, J., Herbert, A. J., & Kelly, A. L. (2022a). Genetic associations with technical capabilities in English academy football players: A preliminary study. *The Journal of Sports Medicine and Physical Fitness*. https://doi.org/10.23736/S0022-4707.22.13945-9

McAuley, A. B. T., Hughes, D. C., Tsaprouni, L. G., Varley, I., Suraci, B., Baker, J., Herbert, A. J., & Kelly, A. L. (2022b). Genetic associations with personality and mental toughness profiles of English academy football players: An exploratory study. *Psychology of Sport and Exercise*, *61*, 102209. https://doi.org/10.1016/j.psychsport.2022.102209

McAuley, A. B. T., Hughes, D. C., Tsaprouni, L. G., Varley, I., Suraci, B., Roos, T. R., Herbert, A. J., & Kelly, A. L. (2021a). Genetic association research in football: A systematic review. *European Journal of Sport Science*, *21*(5), 714–752. https://doi.org/10.1080/17461391.2020.1776401

McAuley, A. B. T., Hughes, D. C., Tsaprouni, L. G., Varley, I., Suraci, B., Roos, T. R., Herbert, A. J., & Kelly, A. L. (2021b). The association of the *ACTN3* R577X and *ACE* I/D polymorphisms with athlete status in football: A systematic review and meta-analysis.

Journal of Sports Sciences, *39*(2), 200–211. https://doi.org/10.1080/02640414.2020.1812195

McAuley, A. B. T., Hughes, D. C., Tsaprouni, L. G., Varley, I., Suraci, B., Roos, T. R., Herbert, A. J., & Kelly, A. L. (2022e). Genetic testing in professional football: Perspectives of key stakeholders. *Journal of Science in Sport and Exercise*, *4*(1), 49–59. https://doi.org/10.1007/s42978-021-00131-3

McAuley, A. B. T., Hughes, D. C., Tsaprouni, L. G., Varley, I., Suraci, B., Roos, T. R., Herbert, A. J., Jackson, D. T., & Kelly, A. L. (2022d). A systematic review of the genetic predisposition to injury in football. *Journal of Science in Sport and Exercise*. Advance online publication. https://doi.org/10.1007/s42978-022-00187-9

McAuley, A. B. T., Hughes, D. C., Tsaprouni, L. G., Varley, I., Suraci, B., Baker, J., Herbert, A. J., & Kelly, A. L. (2022c). Genetic variations between youth and professional development phase English academy football players. *Genes*, *13*(11), 2001. https://doi.org/10.3390/genes13112001

McAuley, A. B. T., Baker, J., Johnston, K., Varley, I., Herbert, A. J., Suraci, B., Hughes, D. C., Tsaprouni, L. G., & Kelly, A. L. (2023). Talent inclusion and genetic testing in sport: A practitioner's guide. *Current Issues in Sport Science*, *8*(1), Article 008. https://doi.org/10.36950/2023.1ciss008

McAuley, A. B. T., Hughes, D. C., Tsaprouni, L. G., Varley, I., Suraci, B., Baker, J., Herbert, A. J., & Kelly, A. L. (in press). Genetic associations with acceleration, change of direction, jump height, and speed in English academy football players. *The Journal of Strength and Conditioning Research*.

Conference proceedings

McAuley, A.B.T., Hughes, D.C., Tsaprouni, L.G., Varley, I., Suraci, B., Roos, T.R., Herbert, A.J., Daniel T. Jackson., & Kelly, A.L. (2021, September). *A systematic review of the genetic predisposition to injury in football*. Paper presented at the 26th Annual Congress of the European College of Sport Science (ECSS). Abstract

McAuley, A.B.T., Hughes, D.C., Tsaprouni, L.G., Varley, I., Suraci, B., Roos, T.R., Herbert, A.J., & Kelly, A.L. (2021, May). *Genetic testing in professional football: Perspectives of key stakeholders*. Paper presented at the 9th annual Conference of the Expertise and Skill Acquisition Network (ESAN), online. Abstract

McAuley, A.B.T., Hughes, D.C., Tsaprouni, L.G., Varley, I., Suraci, B., Roos, T.R., Herbert, A.J., & Kelly, A.L. (2020, December). *Soccer genomics: past, present, and future*. Paper presented at the RESFEST Annual Conference, Birmingham City University, Birmingham.

McAuley, A.B.T., Hughes, D.C., Tsaprouni, L.G., Varley, I., Suraci, B., Roos, T.R., Herbert, A.J., & Kelly, A.L. (2020, November). *The association of the ACTN3 R577X and ACE I/D polymorphisms with athlete status in football: A systematic review and meta-analysis*. Paper presented at the Exercise & Sport Science Australia (ESSA) Twitter Conference.

McAuley, A.B.T., Hughes, D.C., Tsaprouni, L.G., Varley, I., Suraci, B., Roos, T.R., Herbert, A.J., & Kelly, A.L. (2020, October). *Genetic association research in football: A systematic review*. Paper presented at the 25th Annual Congress of the European College of Sport Science (ECSS). Abstract

McAuley, A.B.T., Hughes, D.C., Tsaprouni, L.G., Varley, I., Suraci, B., Roos, T.R., Herbert, A.J., & Kelly, A.L. (2020, August). *The Football Gene Project: A multi-disciplinary investigation into the association of genetic polymorphisms with performance-related phenotypes in football*. Paper presented at the Centre of Life and Sport Sciences (C-LaSS) Research Conference, Birmingham City University, Birmingham.

Acknowledgements

Firstly, I must thank my supervisory team, colleagues, and external collaborators who contributed to the work completed within this thesis. Namely, Dr David Hughes, Dr Loukia Tsaprouni, Dr Ian Varley, Bruce Suraci, Thomas Roos, Dr Adam Herbert, Daniel Jackson, and Professor Joe Baker. Their involvement and wide-ranging areas of expertise ensured this project was completed to the highest possible standard.

However, special mention must go to Dr Adam Kelly for his continual involvement, extensive support, and the opportunities he provided throughout the project. The work ethic, passion, and drive he displayed motivated me to perform to the best of my ability, and my development was significantly enhanced by his valuable knowledge and experience.

It is also important to thank all of the participants who agreed to take part in this research, and the coaches, practitioners, and organisations for providing me with access to their data. Moreover, I am grateful to the Life Sciences department of Birmingham City University for believing in this project and giving me the chance to follow my passion.

Lastly, I wish to thank all my friends, family, and partner Lindsey for supporting me during not only this project but my entire academic journey to this point. They have provided a positive and productive environment and I would not be here today without their help and encouragement, for which I will always be eternally grateful.

This project is dedicated with love to my father Brian and in loving memory of my mother, Tracy, and grandmother Elanor.

Contents

List of Tables

List of Figures

1 General introduction

McAuley, A. B. T., Baker, J., & Kelly, A. L. (2021). How nature and nurture conspire to influence athletic success. In A. L. Kelly, J. Côté, M. Jeffreys, & J. Turnnidge (Eds.), *Birth advantages and relative age effects in sport: Exploring organizational structures and creating appropriate settings* (pp. 159–183). Routledge.

1.1 Introduction

The central determinants of athletic performance have been a main area of investigation within sport science research. It is an enormously complex topic, as both an individual's genetic architecture and environmental exposures require equal deliberation (Davids & Baker, 2007). It has been consistently demonstrated that participation in and commitment to appropriately designed training programmes can result in increases in performance (Kraemer & Ratamess, 2004). However, even complete adherence to an optimally planned training programme may not always lead to expertise (Bouchard, 2012). In contrast, there are individuals who have produced exceptional levels of performance without significant involvement in extrinsic training interventions (Tucker & Collins, 2012). Importantly, adaptations to equivalent training approaches can also vary significantly between individuals (Bouchard, 2012). This unique and large inter-individual variation has an extensive research history. Specifically, the influence of inherited genetic traits versus environmental factors, more commonly known as the 'nature vs. nurture' debate. However, most contemporary researchers now agree that the previous dichotomous argument of nature or nurture is irrelevant, as both an individual's inherited genetic material and their environmental exposures are inevitably jointly responsible for inter-

individual performance disparities (Davids & Baker, 2007; Georgiades et al., 2017; Hambrick et al., 2018; Macnamara et al., 2016; Tucker & Collins, 2012; Ullén et al., 2016; Yan et al., 2016).

Heritability studies have provided strong evidence of a significant heritable component to the various parameters underpinning athletic performance, even when adjusting for environmental effects (Yan et al., 2016). The 'heritability' statistic was introduced with the purpose of measuring the proportion of overall phenotypic variance explained by inter-individual genetic variance in specific environmental contexts (Fisher, 1918). For instance, anthropometric studies produced relatively high heritability estimates for height (~80%), mesomorphy (~80%), skeletal muscle mass (~80%), body mass (~60%), and body mass index (~60%) (Livshits et al., 2016; Peeters et al., 2007; Silventoinen et al., 2008; Visscher et al., 2006). In physiological research, heritability studies have often focused on investigating maximal endurance, strength, and power capacities. Recent meta-analyses have amalgamated the results of these studies and reported relatively high weighted heritability estimates of 72% for VO_2 max (Schutte et al., 2016), as well as 52% for strength and power measurements (Zempo et al., 2017). Similar estimates have been reported for mental traits such as motor control (~70%), motor learning (~70%), cognitive abilities (~50%), and personality dimensions (~50%) (Horsburgh et al., 2009; Missitzi et al., 2013; Pellicciari et al., 2009). A heritability estimate of 66% was also found for overall athletic status, irrespective of the sport (De Moor et al., 2007). In addition, the most comprehensive heritability meta-analysis conducted to date, which comprised ~14 million twin pairs and 17,804 human traits, produced a weighted heritability estimate of 49% across all traits (Polderman et al., 2015). Moreover, no trait produced a weighted heritability estimate of 0%, indicating that all human traits are influenced by heritable factors to some extent (Polderman et al., 2015).

Whilst heritability studies establish a foundation of the potential genetic influence on human traits, they do not reveal which specific biological variants directly contribute to the observed inter-individual differences (Georgiades et al., 2017). However, more sophisticated molecular biology techniques developed over the past two decades have enabled the analysis of specific genetic variants and their association with selected phenotypes (Pitsiladis & Wang, 2011). The focus of current genetic research is on further understanding these genotype-phenotype relationships using genetic association approaches (Guilherme et al., 2014; Visscher et al., 2012; Williams & Folland, 2008). The most common experimental approach employed to identify genetic associations is the candidate gene association study (CGAS; Guilherme et al., 2014). This design uses preselected (i.e., candidate) genetic variants according to their known or postulated biological function and previous results with a relevant trait in a particular cohort (Attia et al., 2009). Three main types of investigations have been performed: (a) *case-control,* which compare the genotype/allele frequency between categorical variables (e.g., athletes and controls; Gineviciene et al., 2014; Juffer et al. 2009; Santiago et al., 2008), (b) *cross-sectional,* which compare genotype/allele associations with quantitative variables (e.g., sprint time and jump height; Massidda et al., 2012; Micheli et al., 2011; Pimenta et al., 2013), and (c) *longitudinal,* which compare genotype/allele associations with responses to specific interventions (e.g., resistance training and aerobic training modifications; Jones et al., 2016; Pickering et al., 2018; Pimenta et al., 2012).

The positives of CGAS include its inexpensiveness and relatively ease to conduct (Guilherme et al., 2014). However, the major limitation is the limited number of genetic variants that can be assessed in each study. This limitation has led to new experimental approaches, with the genome-wide association study (GWAS) becoming one of the most common (e.g., Al-Khelaifi et al., 2019; Pickering et al., 2019; Rodas et al., 2019). A GWAS

does not involve the pre-selection of genetic variants based upon contemporary theoretical suggestions regarding their possible influence on particular traits (Attia et al., 2009). Instead, a GWAS is 'hypothesis-free', so it can analyse an extremely large number (i.e., >1,000,000) of genetic polymorphisms (i.e., variants with a frequency above 1% in the population) and suggest genotype-phenotype associations based solely on observed data (Visscher et al., 2012). This makes the GWAS a more robust genetic association tool as it increases the chance of finding novel genetic variants associated with investigated traits (Bouchard et al., 2011). However, a GWAS is more expensive and requires large homogenous samples in order to reach adequate statistical power (i.e., 5×10^{-8}), due to the number of multiple comparisons being performed (McCarthy et al., 2008). This can be problematic, since high-performance cohorts in sport are very small and heterogenic by nature (Hughes et al., 2011).

Another popular approach is that of the total genotype score (TGS). Analysis of genetic variants in isolation is useful, but also limited within a sporting context as athletic performance is polygenic (i.e., influenced by numerous genes and variants) (Flueck et al., 2010; Hoppeler et al., 2011). The combination of a range of genetic variants is known as a polygenic profile, which Williams and Folland (2008) proposed could be assessed using a simple mathematical equation. To be specific, a TGS is created by assigning each genotype of a genetic variant with a score between 0-2 based on its association with the phenotype. These genotype scores are subsequently summed and transformed into a 0-100 scale by dividing the total score by the maximum possible score and multiplying by 100. Therefore, a TGS assumes genetic variants have a dose-response effect on the phenotype under investigation (Eynon et al., 2011). The sensitivity of a TGS improves as the number of genetic variants included in the polygenic profile increases (Lucia et al., 2010). However, the chances of an individual possessing every favourable allele decreases exponentially as more genetic variants are added (Williams &

Folland, 2008). Hence, one of the limitations of this approach is that all genetic variants are viewed as having the same effect on a phenotype (Eynon et al., 2011). To counteract this, a weighting algorithm can be used, whereby each genetic variant has a multiplier attached that varies depending upon the strength of their association or supporting evidence. For example, Lall et al. (2017) showed that the hazard for incident type 2 diabetes was 3.45 times higher in the highest quintile compared with the lowest quintile using a TGS with a weighted algorithm, after adjusting for body mass index and other known predictors. Therefore, including more genetic variants in a TGS with an appropriate weighting algorithm may improve statistical power (Monnerat-Cahli et al., 2017).

The similarity and frequency of genetic variations are not universal across all population groups. Genetic variation is contingent on a number of factors such as geographical ancestry (i.e., ethnicity) and sex. As such, depending on the characteristics that define particular cohorts, distinct genetic associations may emerge if the allele frequency of a specific genetic variant differs between population substructures (e.g., population stratification; Attia et al., 2009). If a significant association is found in a specific population, replication is required in other populations to substantiate findings and identify the underpinning biological mechanisms (Guilherme et al., 2014). A failure to replicate associations may also indicate that methodological limitations were present in the original study (Ioannidis et al., 2001). For instance, polymorphisms generally have small effects (i.e., odds ratios [OR] of ~1.2 and R^2 of ~1%) on complex traits (i.e., athletic performance; Bouchard, 2011; Tanisawa et al., 2020). Thus, studies that are not sufficiently powered with an adequate sample size and comprise athlete cohorts heterogenous in characteristics influencing associations with performance (e.g., sex, ethnicity, sport, and playing level) may produce false positive and/or negative results.

Genetic association studies have grown extensively in sport since the first polymorphism associated with athletic performance (i.e., *ACE* I/D) was discovered in 1998 (Gayagay et al., 1998; Montgomery et al., 1998). Indeed, a literature review reported that, from the period of 1998-2015, a total of 155 polymorphisms associated with athlete status in sport had been identified (Ahmetov et al., 2016). Interestingly, 77% were identified between 2010-2015, which showcased a remarkable rise of interest within this field of research. Most of this research has investigated genetic associations with athletes from individual sports such as sprinting, long-distance running, cycling, rowing, and swimming (Ahmetov et al., 2016). As such, 93 and 62 genetic polymorphisms were associated with 'endurance' and 'strength/power' sports at least once, respectively. However, only 17 and 14 genetic polymorphisms had their associations replicated in at least two independent cohorts of endurance and strength/power athletes, respectively. These numbers were reduced even further to five and seven respectively, when the criterion was at least three replicated associations in independent cohorts. As such, it is likely that as many as 128, if not more, of these genetic polymorphisms may have had false positive associations and still require further replicatory research (Ahmetov et al., 2016).

Whilst genetic research has predominately focused on individual sports, some recent research has investigated team-sport athletes (e.g., football, rugby, basketball, and volleyball players). Team-sport studies have taken a variety of forms, with some studies having included both team-sport and individual athletes in their population sample (e.g., Orysiak et al., 2015; Durmic et al., 2017), as well as combining groups of different team-sport athletes together (e.g., Ahmetov et al., 2013; Eynon et al., 2014). These studies have limited findings due to the contrasting anthropometric and physiological characteristics between athletes of different sports. Indeed, these differences likely manifest at the molecular level, which weakens the sensitivity of identifying valid genetic associations (Guilherme et al., 2014). Thus, a key issue

within team-sport research is the heterogeneity of population samples. Some studies have been conducted specifically with athletes competing in one team-sport such as football (Santiago et al., 2008), rugby union (Bell et al., 2012), basketball (Garatachea et al., 2014), and volleyball (Ruiz et al., 2011). An important factor to consider in single team-sport studies with regards to senior athletes is inter-positional variations in match demands. Indeed, genetic associations with athlete status in team-sports can greatly depend on the number of players included relative to their on-field positions in sports such as rugby or football (Heffernan et al., 2016; Massidda et al., 2018).

Football appears to be one of the most frequently studied team-sports in sports genomic research. Football is classified as an intermittent sport, with an activity profile comprising numerous variations of movement (e.g., accelerations, changes of direction, dribbling, heading, kicking, jumping, tackling) that fluctuate in frequency depending on factors such as competition level, on-field position, style of play, and tactics employed (Aziz et al., 2008; Bloomfield et al., 2007; Di Salvo et al., 2007). Overall, athletic performance in football is underpinned by a number of multidimensional performer constraints which include anthropometric (e.g., height, body mass), physiological (e.g., endurance, speed), psychological (e.g., personality dimensions, mental toughness), technical (e.g., passing, shooting), and tactical (e.g., perceptual-cognitive expertise, decision-making) capacities (Bergkamp et al., 2019; Dodd & Newans, 2018; Gledhill et al., 2017; Haugen et al., 2014; Murr et al., 2018; Sarmento et al., 2018; Slimani & Nikolaidis, 2019). From an injury perspective, it is estimated that each player will likely sustain two injuries per season that affect their availability for match selection (Ekstrand et al., 2011), with a recent meta-analysis in professional football reporting an injury incidence rate of 8.1/1000 h of exposure (López-Valenciano et al., 2020).

Corresponding with genetic association research in other sports, football genomic studies appear to have mainly focused on case-control CGASs and have reported many contrasting findings. For example, one of the most studied genetic variants in football appears to be the *ACTN3* (rs1815739) polymorphism. Santiago et al. (2008) reported that the *ACTN3* R/R genotype was displayed at a significantly higher frequency in 60 elite Spanish footballers (48.3%) than 102 elite Spanish endurance athletes (26.5%) and 123 controls (28.5%). Subsequently, similar findings have been shown in Russian (Egorova et al., 2014), Turkish (Ulucan et al., 2015), and Iranian (Honarpour et al., 2017) populations. However, Massidda et al. (2014) found there were no significant differences in genotype frequency between 90 elite Italian footballers and 180 controls. Additionally, Cocci et al. (2019) found that the X allele was overrepresented (56.4%) compared to the R allele (43.6%) in 53 elite Italian footballers. Similar contrasting findings have also been reported in studies investigating other genetic variants such as the *ACE* I/D (e.g., Oh et al., 2007; Juffer et al., 2009; Micheli et al., 2011; Gineviciene et al., 2014) and *PPARA* (rs4253778) (e.g., Ahmetov et al., 2013; Proia et al., 2014) polymorphisms. Corresponding to other team-sports, the heterogeneity and size of population samples, as well as the analysis of players as a team instead of on-field positions seem to be common methodological issues (Egorova et al., 2014; Massidda et al., 2018). Thus, currently no single gene or polymorphism appears to be irrefutably associated with performance in football. Although, studies employing TGS designs have found more consistent associations with athlete status and quantitative phenotypes in football (e.g., Egorova et al., 2014; Massidda et al., 2014; Jones et al., 2016; Pickering et al., 2017; 2018). This suggests athletic performance in football may be underpinned by the additive effect of multiple genetic variants. However, at present it is still unclear which genetic variants and polygenic profiles may influence phenotypes in football-specific contexts. Therefore, further examination of genetic associations in football is warranted.

1.2 Philosophical position

It is important that a researcher's philosophical position is explicitly communicated to the reader as it reveals what assumptions the researcher is making regarding their research, which affects the choices applied to the overarching methodology, study designs, data analyses and interpretation of results. A pragmatic philosophical position was adopted during the preparation of this thesis. As a research philosophy, pragmatism embraces a variety of methodological approaches (Maxcy, 2003). Indeed, the philosophical foundation of pragmatism is scientists should employ the methodological approach that is most appropriate to answer specific research questions (Tashakkori & Teddlie, 1998). Pragmatism has been regularly associated with mixed-methods or multimethodology research, wherein the implications of the study are of more importance than following a particular philosophy (Biesta, 2010; Creswell & Plano Clark, 2011; Johnson & Onwuegbuzie, 2004; Maxcy, 2003; Morgan, 2014; Teddlie & Tashakkori, 2009). Pragmatism rejects the overly simplistic objective and subjective philosophical dualism (Biesta, 2010), and enables a pragmatic researcher to resist the artificial dichotomies of positivism and interpretivism (Creswell & Plano Clark, 2011). Instead of viewing positivism and interpretivism as two independent ontological and epistemological perspectives, pragmatism encourages the use of approaches underlying either or both perspectives where applicable (Morgan, 2014).

Whilst positivists claim an objective knowledge attained by hypothesis testing and empirical evidence, and interpretivists suggest reality is more complex and knowledge is relative, pragmatists view the acquisition of knowledge as an objective and subjective continuum instead of two unconnected constructs and are positioned near the centre (Goles & Hirschheim, 2000). Indeed, as positivists typically use quantitative analysis alongside deductive reasoning, and interpretivists normally analyse qualitative data with inductive

reasoning, pragmatists employ a flexible approach by selecting the research design and analysis most befitting the research question (Feilzer, 2010; Morgan, 2007; Pansiri, 2005). As such, pragmatism is often associated with abductive reasoning, whereby the researcher can freely shift between deduction and induction, as well as use a single method, multiple methods, or a mix of methods (Feilzer, 2010; Goldkuhl, 2012; Morgan, 2007). Therefore, adopting a pragmatist philosophical position prioritised the research questions of each chapter and enabled the implementation of the required or optimal methods of inquiry, irrespective of their orientation in traditional philosophical doctrine.

1.3 Thesis aim and project structure

The overarching aim of this thesis was to investigate the association of genetic polymorphisms with phenotypes in football-specific contexts. This project is structured in an article-based format, whereby the forthcoming chapters comprise peer-reviewed journal publications.

The first section is comprised of three research synthesis articles:

(a) Genetic association research in football: A systematic review (Chapter 2). The purpose of this study was to synthesise genetic association studies involving football players to assess the current progress and methodological rigor of this research base, as well as identify limitations and outline future research avenues.

(b) The association of the *ACTN3* R577X and *ACE* I/D polymorphisms with athlete status in football: A systematic review and meta-analysis (Chapter 3). The purpose of this study was to assess the association of the *ACTN3* R577X and *ACE* I/D polymorphisms with athlete status in football via a meta-analysis.

(c) A systematic review of the genetic predisposition to injury in football (Chapter 4). The purpose of this study was to synthesise genetic association studies investigating injury

involving football players to identify which genetic variants have the most empirical evidence.

The second section is comprised of five experimental articles:

(a) Genetic testing in professional football: Perspectives of key stakeholders (Chapter 5). The purpose of this study was to assess the current practical application of genetic testing in professional football and provide an insight into the perspectives of key stakeholders (i.e., coaches, practitioners, players).

(b) Genetic associations with technical capabilities in English academy football players: A preliminary study (Chapter 6). The purpose of this study was to examine the association of eight polymorphisms, both individually and collectively, with dribbling, passing, and shooting in youth male football players.

(c) Genetic associations with personality and mental toughness profiles of English academy football players: An exploratory study (Chapter 7). The purpose of this study was to examine the association of ten polymorphisms, both individually and collectively, with personality dimensions and mental toughness in youth male football players.

(d) Genetic associations with acceleration, change of direction, jump height, and speed in English academy football players (Chapter 8). The purpose of this study was to examine the association of twenty-two polymorphisms, both individually and collectively, with acceleration, change of direction, jump height, and speed in youth male football players.

(e) Genetic variations between youth and professional development phase English academy football players (Chapter 9). The purpose of this study was to examine differences in the genotype frequency distribution of thirty-three polymorphisms, both

individually and collectively, between youth male football players of different age-specific phases.

The final section (Chapter 10) provides an integrative discussion, whereby the key points of the individual articles are summarised and amalgamated to offer a cohesive interpretation of this project's findings and highlight novel contributions to the research field. Limitations of the project, proposed directions for future research, and implications for practice are also outlined before concluding.

1.4 Contextual information

The Football Gene Project is a collaboration between four professional football clubs in England and Birmingham City University to investigate genetic associations with phenotypes in football. The role of the partnered clubs was to collect and supply a wide-range of phenotypic and genotypic de-identified data that the author would then be granted access to for analysis purposes. The phenotypic data collected by the respective clubs were those already routinely measured and administered by appropriately qualified practitioners. The dataset encompassed a total of 177 under-12 to under-23 male academy football players containing the genetic information of each player, alongside 49 anthropometrical, environmental, physiological, psychological, sociological, technical, and tactical phenotypes. Chapters six to nine are the direct analyses of these data, with the phenotypes chosen based on the gaps identified in the literature from the research synthesis section and following extensive cleaning of the dataset.

2 Genetic association research in football: a systematic review

McAuley, A. B. T., Hughes, D. C., Tsaprouni, L. G., Varley, I., Suraci, B., Roos, T. R., Herbert, A. J., & Kelly, A. L. (2021a). Genetic association research in football: A systematic review. *European Journal of Sport Science*, *21*(5), 714–752. https://doi.org/10.1080/17461391.2020.1776401

Abstract

Genetic variation is responsible for a large amount of the inter-individual performance disparities seen in sport. As such, in the last ten years genetic association studies have become more common across a range of sports, including football. However, the progress and methodological rigor of genetic association research in football is yet to be evaluated. Therefore, the aim of this paper was to identify and evaluate all genetic association studies involving football players and outline where and how future research should be directed. Firstly, a systematic search was conducted in the Pubmed and SPORTDiscus databases, which identified 80 eligible studies. Progression analysis revealed that 103 distinct genes have been investigated across multiple disciplines; however, research has predominately focused on the association of the *ACTN3* or *ACE* gene. Furthermore, 55% of the total studies have been published within the last four years; showcasing that genetic association research in football is increasing at a substantial rate. However, there are several methodological inconsistencies which hinder research implications, such as; inadequate description or omission of ethnicity and on-field positions. Furthermore, there is a limited amount of research on several key areas crucial to footballing performance, in particular; psychological related traits. Moving forward,

improved research designs, larger sample sizes, and the utilisation of genome-wide and polygenic profiling approaches are recommended. Finally, we introduce the Football Gene Project, which aims to address several of these limitations and ultimately facilitate greater individualised athlete development within football.

2.1 Introduction

The major determinants of athletic potential is a key topic in the sport science domain. It is widely acknowledged that the dedication to a well-designed training programme will result in a performance increase (Tucker & Collins, 2012). However, dedication to a well-designed training programme does not guarantee elite status (Bouchard, 2012). Some individuals may display high performance levels without substantial extrinsic training interventions (Tucker & Collins, 2012). Furthermore, individuals vary in their response to equivalent training approaches (Bouchard, 2012). Recent research has shown considerable evidence of a significant association between genetics and performance. For instance, heritability studies have now shown that cognitive abilities, motor attributes, morphological dimensions, functional capacities, and personality traits are moderately to highly hereditary (Georgiades et al., 2017). Indeed, in a large cohort study of 4,488 female participants it was shown that genetics were 65.5% responsible for differences in athlete status (De Moor et al., 2007). However, while heritability studies are important, they fail to provide specific information concerning which genes and polymorphisms are responsible for the variations in athletic performance (Guilherme et al., 2014). Therefore, the focus of current genetic research is on further understanding genotype-phenotype relationships.

Three of the most popular approaches for evaluating genotype-phenotype relationships are candidate gene association studies (CGAS), Genome Wide Association Studies (GWAS), and Total Genotype Scores (TGS). A CGAS is defined by selecting a single nucleotide polymorphism (SNP) in a gene and identifying if it is associated with a predefined outcome measure (Guilherme et al., 2014). A GWAS on the other hand, analyses the entire genome without a preceding hypothesis regarding the potential outcomes of genetic variants (Visscher et al., 2012). Hence, a GWAS can identify a substantial number of novel SNPs associated with

a predefined outcome measure (Visscher et al., 2012). In comparison, TGS studies investigate the combined contribution of SNPs on a predefined outcome measure, through the creation and utilisation of a polygenic profile (Williams & Folland, 2008). While all approaches have their own individual weaknesses, each approach contributes crucial pieces of evidence which enhance our understanding of genotype-phenotype associations in sport (Rankinen & Bouchard). Indeed, these approaches have discovered that genotype frequency can be influenced by sex and ethnicity (Ahmetov et al., 2013; Massidda et al., 2018; Yang et al., 2003). Furthermore, sample size significantly influences the potential to discover a significant difference in and between groups, when small differences are being observed (Eynon et al., 2011). Thus, genetic association studies should consider population stratification when recruiting participants and sample size when interpreting results (Little et al., 2009).

The most researched sports within genetic association research are individual sports, such as sprinting, long-distance running, cycling, rowing, and swimming (Ahmetov et al., 2016). However, there has been a recent increase in research interest on team-sports, such as football. Football is classified as an intermittent sport, consisting of rapid changes of speed and movement. Furthermore, each on-field position has a unique physiological demand. For example, forwards can complete up to twice as many sprints as midfielders and defenders per game (Massidda et al., 2018). Overall, football is characterised by several physiological (i.e., aerobic/anaerobic capacity, strength, power, speed, repeated sprint ability [RSA], agility), psychological (i.e., decision-making, anticipation, confidence, fear of failure), and technical (i.e., dribbling, passing, shooting) factors (Murr et al., 2018; Sarmento et al., 2018). Furthermore, the optimisation of these factors is greatly mediated by injury susceptibility (López-Valenciano et al., 2020). Thus, football is a sport which relies on a number or interconnected parameters in order to be successful. Consequently, numerous football genomic

studies have attempted to identify genetic variants associated with these and other factors. However, the full extent of genetic research in football remains unclear. One of the earliest genetic association papers including footballers was conducted two decades ago (Fatini et al., 2000). Twenty years later, it appears there is yet to be a review that fully encompasses all genetic association research on football. As such, the progress and methodological rigor of genetic association research in football is yet to be fully evaluated. Therefore, this paper's aim was to locate all genetic association studies involving footballers and correct this omission.

2.2 Methodology

2.2.1 Search strategy and eligibility criteria

In accordance to the Preferred Reporting Items for Systematic reviews and Meta-Analysis (PRISMA) guidelines (Moher et al., 2009), the following search strategy was implemented. A comprehensive search of the Pubmed and SPORTDiscus databases was conducted on September 12th 2019, using the Boolean search of: (((football OR soccer)) AND ((gene) OR (genetics) OR (polymorphism) OR (genotype) OR (snp) OR (phenotype))). To ascertain if football players were included, studies were required to specify the sport of the athletes involved. Hence, any study that did not specify the sports of the athletes involved were excluded. Specifically, all primary genetic association studies published in English, involving football players, were included.

2.2.2 Data extraction and analysis

The key themes analysed were the progress and methodological rigor of football genetic association studies. Therefore, the year of studies and the number of studies within each year were analysed. Additionally, the type of study (i.e., status, physiological-phenotype, psychological, injury, health, bone-phenotype, career progression) was examined, along with the number of genes included in each study and year. This allowed the progress and growth of

football genetic association studies to be assessed. Secondly, sub themes such as; participant sex, age, ethnicity, status, and playing position were analysed, in order to assess if potential limitations exist. Finally, study designs and the types of athletes involved (individual, team-sport, football) were investigated to support methodological assessment.

2.2.3 Methodological assessment

In order to comprehensively assess the methodological rigor of the genetic association studies included in this review, the STrengthening the REporting of Genetic Association studies (STREGA) initiative was used (Little et al., 2009). STREGA builds on the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement (Von Elm et al., 2008). The additions concern items that are specifically relevant in genetic association studies, such as; population stratification, genotyping errors, Hardy–Weinberg equilibrium, and genetic variant selection rationale. The STREGA recommendations seek to enhance the quality and transparency of reporting in genetic association studies.

2.3 Results

2.3.1 Search process

The systematic search process (see **Figure 2.1**) initially identified 993 studies. Following the removal of duplicates (122), article titles were then evaluated which resulted in the exclusion of 737 studies. The abstracts of the remaining 134 studies were then assessed, leading to the removal of 43 more studies. Full-text critical methodological assessment of the remaining 91 studies followed, resulting in the exclusion of eleven more studies. This culminated in 80 studies being judged as adequately meeting the predetermined inclusion criteria and subsequently being included in the final analysis.

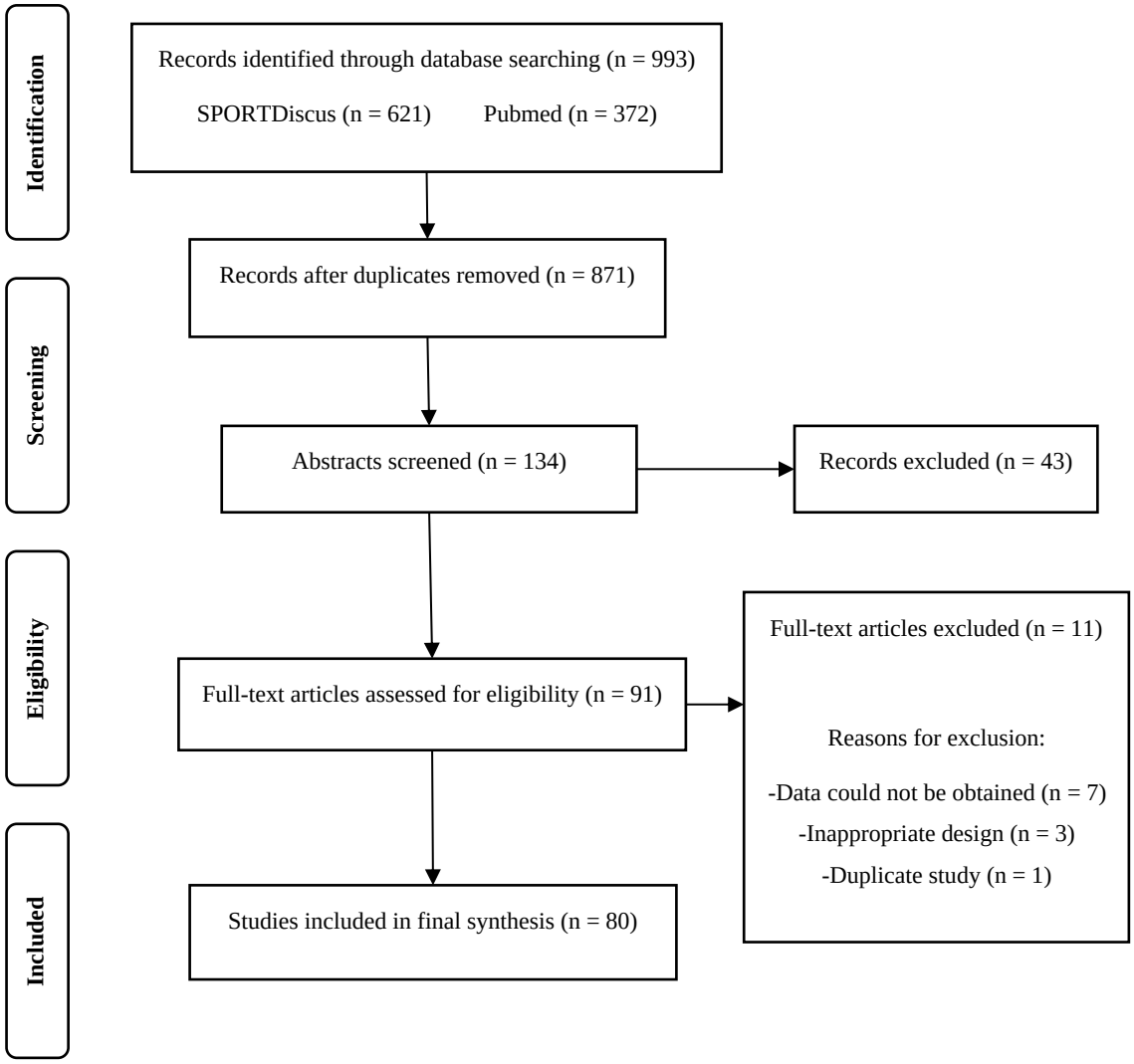


Figure 2.1. Flow chart of systematic search process

2.3.2 Study characteristics

Of the 80 included studies (see **Table 2.1**), there were 66 CGAS, six TGS studies, seven epigenetic studies, and one study which included a GWAS and CGAS design. Fifty studies included only footballers, eight included team-sport athletes (including footballers), and 22 included individual and team-sport athletes (including footballers). Fifty-eight studies focused exclusively on male subjects, three focused on females, and 16 included mixed sexes, whilst the remaining three studies did not include the sex of their subjects. Studies included a range of athlete statuses (elite = 48, sub-elite = 3, amateur = 2, youth = 16) with seven including a

combination of statuses, and four failing to report the players' specific athletic status. Only five studies included the on-field positions of the football players, whilst the remaining 75 did not. The focus of the studies included: Injury (*n* = 22), Athletic Status (*n* = 20), Physiological Phenotype (*n* = 20), Epigenetic (*n* = 7), Health (*n* = 5), Psychological (*n* = 3), Bone-Phenotype (*n* = 2), and Career Progression (*n* = 1). The total number of football players included across all studies was 8,155 (Caucasian = 3754, Mixed = 2061, Unknown = 1562, Japanese = 480, Iranian = 147, Egyptian = 68, Korean = 58, Turkish = 25), ranging from four to 694 player sample sizes, with a median sample size of 60. Ethnicities were studied in isolation (Caucasian = 32, Iranian = 2, Korean = 2, Egyptian = 1, Japanese = 1, Turkish = 1) and mixed (22), while 22 studies did not specify the ethnicity of the players. The age of the players ranged from ten to 71 years old. Finally, 103 distinct genes were included in the studies (see **Figure 2.2** for the names and frequency of genes studied). The most frequently studied genes were actinin alpha 3 (*ACTN3*; *n* = 27) and angiotensin I converting enzyme (*ACE*; *n* = 25).

2.3.3 STREGA adherence

Of the 80 full-text genetic studies that were included, 69 association studies were eligible for STREGA adherence analysis. The remaining 11 were either epigenetic (n = 7) or intervention (n = 4) studies. Overall, the average adherence score of the 69 studies to the STREGA guidelines was 84%. Furthermore, twelve (17%) studies scored between 50-74%, with 26 (38%) studies scoring between 75-89%, and the remaining 31 (45%) scoring between 90-100% adherence.

Table 2.1. Study characteristics and STREGA adherence score

Author/Year	Sample	Sport	Participants					Level	Study			STREGA	Genes
			Ethnicity	Football players	Positions	Sex	Age		Design	Protocol	Type		
Fatini et al. (2000)	28 Italian elite Caucasian male footballers (aged 20.2 ± 4.4 years) vs 155 untrained Caucasian males (aged 23 ± 2.3 years)	Football	Caucasian	28	No	Male	20.2 ± 4.4	Elite	CGAS	Observation	Health	64%	*ACE, AT1R*
Rizzo et al. (2003)	75 Italian male academy footballers (aged 15 ± 1.2 years) vs 52 untrained males (aged 15 ± 1.6 years)	Football	Unknown	75	No	Male	15 ± 1.2	Youth	CGAS	Observation	Health	73%	*ACE*
Oh (2007)	139 Korean elite male athletes (aged 20.7 ± 1.3 years): basketball (n = 15), football (n = 41), baseball (n = 31), gymnastics (n = 12), volleyball (n = 7), long-distance running (n = 8), judo (n = 8), ice hockey (n = 17) vs 163 non-athletes	Mixed	Korean	41	No	Male	20.7 ± 1.3	Elite	CGAS	Observation	Status	68%	*ACE*
Santiago et al. (2008)	60 elite male football players (17-32 years) vs 102 Spanish elite male endurance athletes (19–38 years) vs 123 untrained Spanish males (aged 19–50 years)	Mixed	Mixed	60	No	Male	17-32	Elite	CGAS	Observation	Status	59%	*ACTN3*
Terrel et al. (2008)	195 male college American football players vs 18 male and 18 female college football players (aged 18-30 years).	Team	Unknown	36	No	Mixed	18-30	Youth	CGAS	Observation	Injury	95%	*APOE, MAPT*

Juffer et al. (2009)	54 elite male footballers (age 18–32 years) vs 52 elite Spanish male endurance runners (aged 19–38 years) vs 123 untrained Spanish males (aged 19–50 years)	Mixed	Mixed	54	No	Male	18-32	Elite	CGAS	Observation	Status	91%	*ACE, GDF8, AMPD1*
Diogenes et al. (2010)	46 academy male Brazilian football players (aged 11–14 years)	Football	Mixed	46	No	Male	11-14	Youth	CGAS	Observation	Bone-phenotype	86%	*VDR*
Ginevičienė et al. (2010)	193 elite male (n = 152) and female (n = 41) Lithuanian athletes (aged 22.0 ± 6.3 years): biathlon (n = 5), cross-country skiing (n = 12), road cycling (n = 12), pentathlon (n = 4), swimming (n = 13), rowing (n = 9), track and field (long distance) athletics (n = 9), track and field (short distance) athletics (n = 20), kayaking (n = 13), weightlifting (n = 31), boxing (n = 6), wrestling (n = 10), tennis (n = 3), football (n = 32), handball (n = 14) vs 250 untrained Lithuanian controls (167 males and 83 females aged 36.2 ± 7.2 years)	Mixed	Caucasian	32	No	Mixed	22.0 ± 6.3	Elite	CGAS	Observation	Physiological phenotype	77%	*ACE, ACTN3, PPARGC1A, PPARA*
Kim et al. (2010)	81 male Korean athletes (aged 21.3 ± 1.2 years): long-distance running (n = 8), football (n = 17), baseball (n = 8), basketball (n = 10), volleyball (n = 8), ice hockey (n = 8), judo (n = 8), taekwondo (n = 6), gymnastics (n = 8) vs 33 untrained Korean males (aged 22.2 ± 1.9 years).	Mixed	Korean	17	No	Male	21.3 ± 1.2	Sub-elite	CGAS	Observation	Health	86%	*B3AR*

Tierney et al. (2010)	163 male college American football athletes and 33 college female football athletes (aged 19.7 ± 1.5 years)	Team	Unknown	33	No	Female	19.7 ± 1.5	Youth	CGAS	Observation	Injury	95%	*APOE*
Holdys et al. (2011)	119 trained males and 37 trained females vs 35 untrained males and 48 untrained females. All participants were Caucasian and Polish (aged 18-26 years). Participants came from 23 different sports (including football) and were characterized into 3 groups: speed strength, endurance speed strength and endurance sports.	Mixed	Caucasian	Unknown	No	Mixed	18-26	Mixed	CGAS	Observation	Physiological phenotype	91%	*ACE*
Micheli et al. (2011)	125 medium-high level Italian Caucasian male football players (aged U-17) vs 152 untrained controls	Football	Caucasian	125	No	Male	U-17	Youth	CGAS	Observation	Physiological phenotype	59%	*ACE, VDR*
Sessa et al. (2011)	82 elite male Caucasian Italian athletes, 29 sprinters, short distance swimmers, and volleyball players and 53 football, basketball, and hockey players (aged 24.9 ± 8.5 years) vs 269 untrained Italian males (aged 26.7 ± 2.5 years).	Mixed	Caucasian	Unknown	No	Male	24.9 ± 8.5	Elite	CGAS	Observation	Status	68%	*ACE, ACTN3, NOS3, UCP2, UCP3*

Eynon et al. (2012)	60 elite football players (aged 17-32 years) vs 100 elite endurance athletes (aged 20-39 years) vs 53 elite power athletes (aged 20-33 years) vs 100 non-athletic controls (aged 19-32 years). All participants were Spanish male Caucasians.	Mixed	Caucasian	60	No	Male	17-32	Elite	CGAS	Observation	Status	82%	*NOS3*
Pimenta et al. (2012)	37 elite male footballers from the Brazilian first division	Football	Unknown	37	No	Male	Unknown	Elite	CGAS	Intervention	Physiological phenotype	68%	*ACTN3*
Ahmetov et al. (2013)	665 elite Caucasian male and female Russian athletes in badminton (n = 16), baseball (n = 28), basketball (n = 85), beach volleyball (n = 10), court tennis (n = 33), football (n = 241), futsal (n = 9), handball (n = 24), ice hockey (n = 55), rugby (n = 48), softball (n = 31), table tennis (n = 14), volleyball (n = 53), water polo (n = 18) vs 1706 untrained Russian males and females.	Team	Caucasian	241	No	Mixed	Unknown	Elite	CGAS	Observation	Status	77%	*PPARA*
Ficek et al. (2013)	91 elite male football players (aged 23 ± 3 years) with surgically diagnosed ACL ruptures vs 143 healthy male elite football players (aged 25.2 ± 2.6 years). All participants were Polish East-Europeans.	Football	Caucasian	234	No	Male	23 ± 6	Elite	CGAS	Observation	Injury	95%	*COL1A1*
Pimenta et al. (2013)	200 elite Brazilian footballers (aged 24.4 ± 2.0 years)	Football	Unknown	200	No	Male	24.4 ± 2.0	Elite	CGAS	Observation	Physiological phenotype	68%	*ACTN3*

Pruna et al. (2013)	73 elite male football players (aged 19–35 years) of White, Black-African and Hispanic origin	Football	Mixed	73	No	Male	19–35	Elite	CGAS	Observation	Injury	100%	*ELN, TTN, SOX15, IGF2, CCL2, COL1A1, COL5A1, TNC*
Andreeva et al. (2014)	16 African Zulu female football and netball players (aged 20.8 ± 3.1 years) and 23 Bulgarian Caucasian female football players (aged 22.6 ± 2.0 years). There were control groups of 23 and 42 female students and population cohorts of 104 and 114 subjects respectively	Team	Mixed	32	No	Female	20.8 ± 3.1	Youth	CGAS	Observation	Status	86%	*ACE*
Egorova et al. (2014)	246 male Russian football players - 51 elite (aged 23.9 ± 0.6 years) 81 sub-elite (aged 23.0 ± 0.7 years + 114 non-elite (aged 10.6 ± 0.1 years). elite and sub-elite football players with exact specialisation were classified as goalkeepers (n = 27), attackers (wing-forwards and centre-forwards) (n = 14), defenders (n = 29) and midfielders (n = 13). Controls were 872 untrained males (aged 19.8 ± 0.2 years). All participants were Caucasian.	Football	Caucasian	246	Yes	Male	10-24	Mixed	TGS	Observation	Status	86%	*ACE, ACTN3, PPARA, PPARG, PPARGC1A, PPARD, TFAM, UCP2*
Eynon et al. (2014)	888 male Caucasian elite and national level athletes (305 endurance, 378 sprint/power, 205 team sport) vs 568 untrained controls, from Poland, Russia and Spain.	Mixed	Caucasian	53	No	Male	Unknown	Mixed	CGAS	Observation	Status	91%	*ACTN3*

Gineviciene et al. (2014)	199 Lithuanian sub-elite male footballers (aged 17-20 years) vs 167 untrained men (aged 18-22 years), forwards (n = 44), defenders (n = 63), midfielders (n = 75) and goalkeepers (n = 17). All participants were Caucasian.	Football	Caucasian	199	Yes	Male	17-20	Sub-elite	CGAS	Observation	Status	82%	*ACE, PPARGC1A, PPARA*
Massidda et al. (2014)	90 elite male football players (aged 25.5 ± 6.5 years) vs 180 randomly selected healthy nonathletic Italian males. 12 of the football players were not European, while the remaining 78 were of Italian descent.	Football	Mixed	90	No	Male	25.5 ± 6.5	Elite	TGS	Observation	Physiological phenotype	77%	*ACE, ACTN3, BDKRB2, VDR*
Proia et al. (2014)	60 male elite football players (aged 22.5 ± 2.2 years) vs 30 untrained male students (aged 21.2 ± 2.3 years). All participants were Caucasian.	Football	Caucasian	60	No	Male	22.5 ± 2.2	Elite	CGAS	Observation	Status	91%	*PPARA*
Saber-Ayad et al. (2014)	68 elite male football players (aged 17-21 years) vs 100 untrained male students (aged 17-21 years). All participants were asymptomatic Egyptians	Football	Egyptian	68	No	Male	18.8 ± 1.6	Elite	CGAS	Observation	Health	95%	*ACE*
Alfieri et al. (2015)	5 male Caucasian recreational football players (aged 31.8 ± 5.4 years)	Football	Caucasian	5	No	Male	31.8 ± 5.4	Amateur	N/A	N/A	Epigenetic	N/A	*PPARG, ADIPOQ, AMPKA1, AMPKA2, TFAM, NAMPT, PGC1A, SIRT1*

Atanasov et al. (2015)	52 Bulgarian Caucasian male elite/sub-elite athletes (aged 21.3 ± 1.5 years) vs 109 untrained Caucasian males (aged 20.6 ± 1.9 years). Athletes included, long and triple jump, javelin throw, running: 100m, 200m, 400m, swimming: 200m, 400m, sprint cycling, boxing, wrestling, football, volleyball, handball.	Mixed	Caucasian	Unknown	No	Male	21.3 ± 1.5	Mixed	CGAS	Observation	Physiological phenotype	91%	*ACTN3, AMPD1*
Filonzi et al. (2015)	50 elite athletes (16 females and 34 males) belonging to Caucasian, Afro-American, Afro-European and Maori ethnicities. Sports included football (n = 4), basketball (n = 10), tennis (n = 6), volleyball (n = 6), canoeing (n = 2), rugby (n = 10), baseball (n = 6) and track and field (n = 6) vs 100 (40 females and 60 males age, sex and ethnicity-matched) practicing sport activity at lower levels	Mixed	Mixed	4	No	Mixed	Unknown	Elite	CGAS	Observation	Psychological	82%	*MSTN, 5HTT, SLC6A3, MAOA*
Jeong et al. (2015)	9 males who regularly participated in team sport (aged 25 ± 4 years)	Team	Unknown	Unknown	No	Male	25 ± 4	Amateur	N/A	N/A	Epigenetic	N/A	*PPARGC1A*
Massidda et al. (2015a)	178 Italian elite and sub-elite male athletes in football (n = 64), hockey (n = 10), power (n = 64), endurance (n = 40) vs 190 untrained males	Mixed	Caucasian	64	No	Male	Unknown	Mixed	CGAS	Observation	Status	95%	*ACTN3*
Massidda et al. (2015b)	54 male Caucasian elite football players (aged 25.9 ± 4.3 years)	Football	Caucasian	54	No	Male	25.9 ± 4.3	Elite	CGAS	Observation	Injury	82%	*VDR*

Massidda et al. (2015c)	173 Italian Caucasian male elite football players (aged 19.4 ± 5.2 years)	Football	Caucasian	173	No	Male	19.4 ± 5.2	Elite	CGAS	Observation	Injury	91%	*MCT1*
Pruna et al. (2015)	73 elite male football players (aged 19–35 years) of White, Black-African and Hispanic origin	Football	Mixed	73	No	Male	19-35	Elite	CGAS	Observation	Injury	82%	*ELN, TTN, SOX15, IGF2, CCL2, COL1A1, COL5A1, TNC*
Ulucan et al. (2015)	25 male Turkish professional footballers	Football	Turkish	25	No	Male	Unknown	Elite	CGAS	Observation	Status	59%	*ACE, ACTN3*
Varley et al. (2015)	518 elite athletes (449 male and 69 female). Sports included, football (n = 218), cricket (n = 156), track and field (n = 67, running events n = 62), rowing (n = 13), boxing (n = 2), tennis (n = 12), hockey (n = 26) and gymnastics (n = 7). Athletes were mainly white Caucasian (83.2% in the stress fracture cases and 79.9% in the non-stress fracture controls)	Mixed	Mixed	218	No	Mixed	24.2 ± 5.5	Elite	CGAS	Observation	Injury	95%	*RANK, RANKL, OPG*
Artells et al. (2016)	60 elite male European football players (aged 25.52 ± 2.5 years)	Football	Unknown	60	No	Male	25.5 ± 2.5	Elite	CGAS	Observation	Injury	82%	*ELN*
Bondareva et al. (2016)	28 elite male footballers vs 70 non-athletes	Football	Unknown	28	No	Male	Unknown	Elite	CGAS	Observation	Physiological phenotype	73%	*UCP1, UCP2, UCP3, FTO*
Cięszczyk et al. (2016)	106 elite Polish male footballers - forwards (n = 28), defenders (n = 38), midfielders (n = 32), and goalkeepers (n = 8). Controls consisted of 115 untrained males. All participants were Caucasian.	Football	Caucasian	106	Yes	Male	Unknown	Elite	CGAS	Observation	Status	82%	*ACE*

Coelho et al. (2016)	353 Brazilian male soccer players of different age groups: 43 U-14 (14 ± 0.3 years), 68 U-15 (15 ± 0.4 years), 44 U-17 (16 ± 0.6 years), 115 U-20 (18 ± 0.7 years), 83 Professionals (23 ± 1.7 years) vs 100 Brazilian untrained males (aged 10-14 years).	Football	Mixed	353	No	Male	14-24	Elite	CGAS	Observation	Status	91%	*ACE*
Dinç et al. (2016)	48 elite male footballers in Turkey vs 48 untrained males (aged 18-27 years)	Football	Unknown	48	No	Male	18-27	Elite	CGAS	Observation	Physiological phenotype	59%	*MTHFR*
Gill et al. (2016)	15 - 8 male, 7 female (aged 19.4 ± 1.5 years) National Collegiate Athletic Association Division III athletes with sports related concussion vs 16 non-concussed athletes - 7 male, 9 female (aged 18.5 ± 0.4 years).	Team	Caucasian	4	No	Mixed	19.4 ± 1.5	Youth	GWAS	Observation	Epigenetic	N/A	N/A
Jones et al. (2016)	Study 1 - 28 Caucasian male athletes (aged 18-20 years) squash (n = 1), swimming (n = 7), running (n = 1), ski/snowboard (n = 4), football (n = 1), lacrosse (n = 2), badminton (n = 1), motorsport (n = 1), cycling (n = 4), cricket (n = 2), volleyball (n = 1), fencing (n = 1) and rugby union (n = 2). Study 2 - 39 male football players (aged 16-19 years)	Mixed	Unknown	40	No	Male	16-20	Youth	TGS	Intervention	Physiological phenotype	86%	*ACE, ACTN3, ADRB2, AGT, BDKRB2, COL5A1, CRP, GABPB1, IL6, PPARA, PPARGC1A, TRHR, VDR, VEGFA*
Massidda et al. (2016)	128 elite Italian male football players (aged 16.3 ± 1.3 years)	Football	Caucasian	128	No	Male	15.6 ± 1.8	Elite	CGAS	Observation	Physiological phenotype	82%	*MCT1*
Petito et al. (2016)	133 elite male football, basketball and hockey players.	Team	Unknown	133	No	Male	Unknown	Elite	CGAS	Observation	Psychological	95%	*5HTT*

Wessner et al. (2016)	56 power athletes - sprinters and jumpers (n = 49), throwers (n = 5), weightlifters (n = 2), 86 endurance athletes (middle and long distance runners (n = 63), road cyclists (n = 17), triathletes (n = 5), biathletes (n = 1), 143 team sport athletes (football players (n = 82), handball players (n = 61), and 216 healthy non-athletic controls. Participants included both sex and were all Caucasian (aged 18-83 years).	Mixed	Caucasian	82	No	Mixed	23.5 ± 4.8	Elite	CGAS	Observation	Status	100%	*ACTN3, ADRB1, ADRB2, ADRB3*
Cauci et al. (2017)	60 Italian athletes - 25 females and 35 males (aged 33.9 ± 13.3 years) in swimming (n = 10) football (n = 7) volleyball (n = 7) rugby (n = 6) weight lifting (n = 5) track-and-field sports (n = 5) figure skating (n = 4) artistic gymnastics/competitive dancing (n = 4) basketball (n = 3) triathlon (n = 3) sailing (n = 3) discus throw (n = 2) martial arts (n = 1).	Mixed	Caucasian	7	No	Mixed	33.9 ± 13.3	Unknown	CGAS	Observation	Injury	91%	*VDR*
Cięszczyk et al. (2017)	229 elite Polish football players who suffered an ACL - 158 males (aged 26 ± 4 years) and 71 females (aged 25 ± 4 years) vs 143 uninjured athletes, 99 males (aged 25 ± 3 years) and 44 females (aged 29 ± 2 years). All participants were Caucasian.	Football	Caucasian	229	No	Mixed	26 ± 4	Elite	CGAS	Observation	Injury	95%	*ACAN, BGN, DCN, VEGFA*
Dionísio et al. (2017)	220 elite male footballers (aged 14-20 years) in Brazil	Football	Unknown	220	No	Male	14-20	Elite	CGAS	Observation	Physiological phenotype	91%	*ACTN3, AMPD1, ACE, AGT*

Domańska-Senderowska et al. (2017)	22 football players (aged 17.5 ± 0.7 years)	Football	Unknown	22	No	Unknown	17.5 ± 0.7	Youth	N/A	N/A	Epigenetic	N/A	N/A
Durmic et al. (2017)	107 elite Caucasian male athletes: 17 sprint/power (short-distance runners, swimmers competing in events < 200 m), 36 endurance athletes (rowers, football players, middle distance swimmers), and 54 athletes from mixed sports (water polo, handball and volleyball)	Mixed	Caucasian	Unknown	No	Male	24.7 ± 4.3	Elite	CGAS	Observation	Health	82%	*ACE, ACTN3*
Galeandro et al. (2017)	43 elite male football players vs 128 untrained controls	Football	Mixed	43	No	Male	25 ± 6	Elite	CGAS	Observation	Status	77%	*ACE, ACTN3*
Honarpour et al. (2017)	90 elite male Iranian football players vs 200 unrelated healthy males	Football	Iranian	90	No	Male	Unknown	Elite	CGAS	Observation	Status	82%	*ACTN3*
Mancini et al. (2017)	10 lifelong football-trained men (aged 68.2 ± 3.0 years) and 10 active untrained healthy men (aged 66.7 ± 1.3 years).	Football	Unknown	10	No	Male	68.2 ± 3.0	Unknown	N/A	N/A	Epigenetic	N/A	N/A
Pickering et al. (2017)	18 male college footballer players (aged 16-19 years)	Football	Unknown	18	No	Male	16-19	Youth	TGS	Intervention	Physiological phenotype	N/A	*IL6, CRP, TNF, SOD2, GSTM1, GSTT1*
Pruna et al. (2017)	74 elite male football players (aged 19–35 years) of White, Black-African and Hispanic origin.	Football	Mixed	74	No	Male	19-35	Elite	CGAS	Observation	Injury	77%	*LIF, CCL2, GEFT, MYF5, DES, HGF, MMP3, GDF5*
Żychowska et al. (2017)	9 male football players (aged 19.8 ± 0.6 years) vs 9 untrained males (aged 19.7 ± 0.87 years)	Football	Unknown	9	No	Male	19.8 ± 0.6	Unknown	N/A	N/A	Epigenetic	N/A	N/A

Cochrane et al. (2018)	250 collegiate student-athletes - 66 females, 184 males (aged 19.0 ± 1.3 years). 95 American football, 38 men's baseball, 35 men's football, 32 women's football, 20 women's softball, 16 men's basketball, 2 women's basketball, 1 women's cross-country, and 1 women's track and field	Mixed	Unknown	67	No	Mixed	19.0 ± 1.3	Youth	CGAS	Observation	Psychological	91%	*APOE, COMT, DRD2*
Coelho et al. (2018)	353 elite male Brazilian football players (aged 14-27 years) vs 100 untrained - 50 males, 50 females (aged 8-14 years)	Football	Mixed	353	No	Male	14-27	Elite	CGAS	Observation	Career progression	77%	*ACTN3*
Domańska-Senderowska et al. (2018)	22 male football players (aged 17.5 ± 0.7 years)	Football	Unknown	22	No	Male	17.5 ± 0.7	Unknown	N/A	N/A	Epigenetic	N/A	*PPARD*
Larruskain et al. (2018)	107 Caucasian male football players (aged 20 ± 4 years), 28 players belonged to the First team, 43 to the two Reserves teams, and 36 to the two U-19 teams.	Football	Caucasian	107	No	Male	20 ± 4	Mixed	CGAS	Observation	Injury	91%	*MMP3, COL5A1, MMP1, NOS3, DCN, HIF1A, MMP12, CASP8, ADAM12, SOX15, TNC, COL1A1, CCL2, VEGFA, ADAMTS5, ACTN3, ACAN, ADAMTS2, IL6, GDF5, ACE, COL12A1, SOD2, MLCK, TIMP2, IL6R, ADAMTS14, EMILIN1, CASP8, TTN, IGF2, TNF, IL1A, CCR2, IL1B*

Lulińska-Kuklik et al. (2018)	A total of 134 elite male Polish football players (aged 23.4 ± 3.1 years), with surgically diagnosed primary ACL ruptures vs 211 apparently healthy, male elite football players (aged 25.3 ± 3.4 years), without any self-reported history of ligament or tendon injury.	Football	Caucasian	134	No	Male	23.4 ± 3.1	Elite	CGAS	Observation	Injury	95%	*COL5A1*
Massidda et al. (2018)	A total of 1475 Caucasian males (694 top level football players and 781 controls) from Italy (n = 360), Poland (n = 665), Lithuania (n = 302), Ukraine (n = 136) and Malta (n = 12) participated in the study.	Football	Caucasian	694	Yes	Male	Unknown	Elite	CGAS	Observation	Status	86%	*MCT1*
McCabe and Collins (2018)	289 male football players (aged 18–32 years) including 46 professional, 98 semi-professional and 145 amateur players.	Football	Unknown	289	No	Male	18-32	Mixed	CGAS	Observation	Injury	68%	*GDF5*, *AMPD1*, *COL5A1*, *IGF2*
Pickering et al. (2018)	42 male football players (aged 16–19 years) from a college football academy	Football	Unknown	42	No	Male	16-19	Youth	TGS	Intervention	Physiological phenotype	N/A	*VEGF*, *ADRB2*, *CRP*, *PPARGC1A*
Terrell et al. (2018)	817 college American football males, 155 male and female football, as well 84 male and female basketball, softball, men's wrestling and club men's rugby.	Mixed	Mixed	155	No	Mixed	19.7 ± 1.5	Youth	CGAS	Observation	Injury	95%	*APOE*, *MAPT*, *IL6*, *IL6R*

Varley et al. (2018a)	117 male academy football players. Participants were made up from a variety of ethnicities (64 Caucasian, 19 Caucasian/Black dual heritage, 11 Black Caribbean, 4 Black African and 1 Asian) and were composed of differing playing positions (42 midfielders, 29 defenders, 19 forwards and 9 goalkeepers).	Football	Mixed	117	Yes	Male	>16	Youth	CGAS	Observation	Bone-phenotype	91%	*RANK, RANKL, OPG, WNT, P2X7R, SOST, MP3K, IL6*
Varley et al. (2018b)	518 male (n = 449) and female (n = 69) elite athletes (aged 24.2 ± 5.5 years). Participating elite athletes competed in various sports including, football (n = 218), cricket (n = 156), track and field (n = 67, running events n = 62), rowing (n = 13), boxing (n = 2), tennis (n = 12), hockey (n = 26) and gymnastics (n = 7), with each sport having both stress fracture cases and non-stress fracture control participants. Elite athletes were mainly white Caucasian (83.2% in the stress fracture cases and 79.9% in the non-stress fracture controls).	Mixed	Mixed	218	No	Mixed	24.2 ± 5.5	Elite	CGAS	Observation	Injury	95%	*VDR, WNT, SOST, COL1A1, LRP5, CTR, GC*
Clos et al. (2019)	43 elite male football players (aged 20-37 years) of Caucasian, Black-African and Hispanic origin	Football	Mixed	43	No	Male	20-37	Elite	CGAS	Observation	Injury	91%	*ACTN3*

Cocci et al. (2019)	113 Italian Caucasian males. 37 combat sport athletes (aged 25.9 ± 9.3 years), 21 motorcycle riders (aged 22.5 ± 7.4 years) and 55 football players (aged 24.5 ± 8.7 years).	Mixed	Caucasian	55	No	Male	24.5 ± 8.7	Elite	TGS	Observation	Status	91%	*ACE, ACTN3, PPARA, CKMM*
Falahati and Arazi (2019)	29 elite Iranian male football players vs 28 untrained football players	Football	Iranian	57	No	Male	Unknown	Elite	CGAS	Observation	Physiological phenotype	95%	*ACE*
Jeremic et al. (2019)	27 female football players pertaining to the Serbian national U18 team (16-18 years old).	Football	Unknown	27	No	Female	16-18	Elite	CGAS	Observation	Physiological phenotype	73%	*ACE, ACTN3*
Koku et al. (2019)	100 healthy male Caucasian football players aged 18-30 years vs 101 untrained males at similar age.	Football	Caucasian	100	No	Male	18-30	Sub-elite	CGAS	Observation	Physiological phenotype	77%	*ACTN3*
Kumagai et al. (2019)	1311 elite male (870) and female (441) athletes. Of which 480 were football players (aged 20.6 ± 2.9 years).	Mixed	Japanese	480	No	Mixed	20.6 ± 2.9	Elite	CGAS	Observation	Injury	100%	*ESR1*
La Montagna et al. (2019)	30 elite football players from different nationalities	Football	Mixed	30	No	Unknown	Unknown	Elite	CGAS	Observation	Injury	77%	*ACTN3, COL5A1, MCT1, VEGF, HFE*
Lulińska-Kuklik et al. (2019)	229 elite football players who suffered an ACL injury - 164 males and 65 females (aged 26 ± 4 years) vs 192 controls - 107 males and 85 females (aged 25 ± 3 years).	Football	Caucasian	229	No	Mixed	26 ± 4	Elite	CGAS	Observation	Injury	91%	*TNC*
Massidda et al. (2019)	257 elite male Italian football players (aged 21.± 5.3 years) vs 263 untrained males (aged 22.4 ± 6.2 years)	Football	Caucasian	257	No	Male	21.± 5.3	Elite	CGAS	Observation	Injury	77%	*ACTN3*

Monnerat et al. (2019)	25 elite male football players (aged 25.5 ± 4.3 years) vs 2504 individuals from the populations deposited in the 1000genomes database.	Football	Mixed	25	No	Male	25.5 ± 4.3	Elite	CGAS	Observation	Physiological phenotype	86%	*AMPD1, ACTN3, PPARGC1A, MCT1, COL5A1, CHRM2, MMP3, FTO, CYP1A2*
Pickering et al. (2019)	48 elite Caucasian youth football players (aged 12-18 years)	Football	Caucasian	48	No	Unknown	12-18	Youth	CGAS + GWAS	Observation	Physiological phenotype	95%	*ACE, ACTN3, ADRB2, AGT, AMPD1, CKM, GABRR1, HSD17B14, IGF1, IGF2, IL6, MTHFR, PPARA, PPARG, UCP2*
Stastny et al. (2019)	146 male (n = 90) and female (n = 56) youth players (aged 13–15 years) of basketball (n = 54), soccer (n = 50), and handball (n = 32)	Team	Unknown	146	No	Mixed	13-15	Youth	CGAS	Observation	Physiological phenotype	86%	*COL5A1, GDF5, PPARA*

Note. CGAS = Candidate Gene Association Study; GWAS = Genome Wide Association Study; TGS = Total Genotype Score; ACL = Anterior Cruciate Ligament; U- = Under.

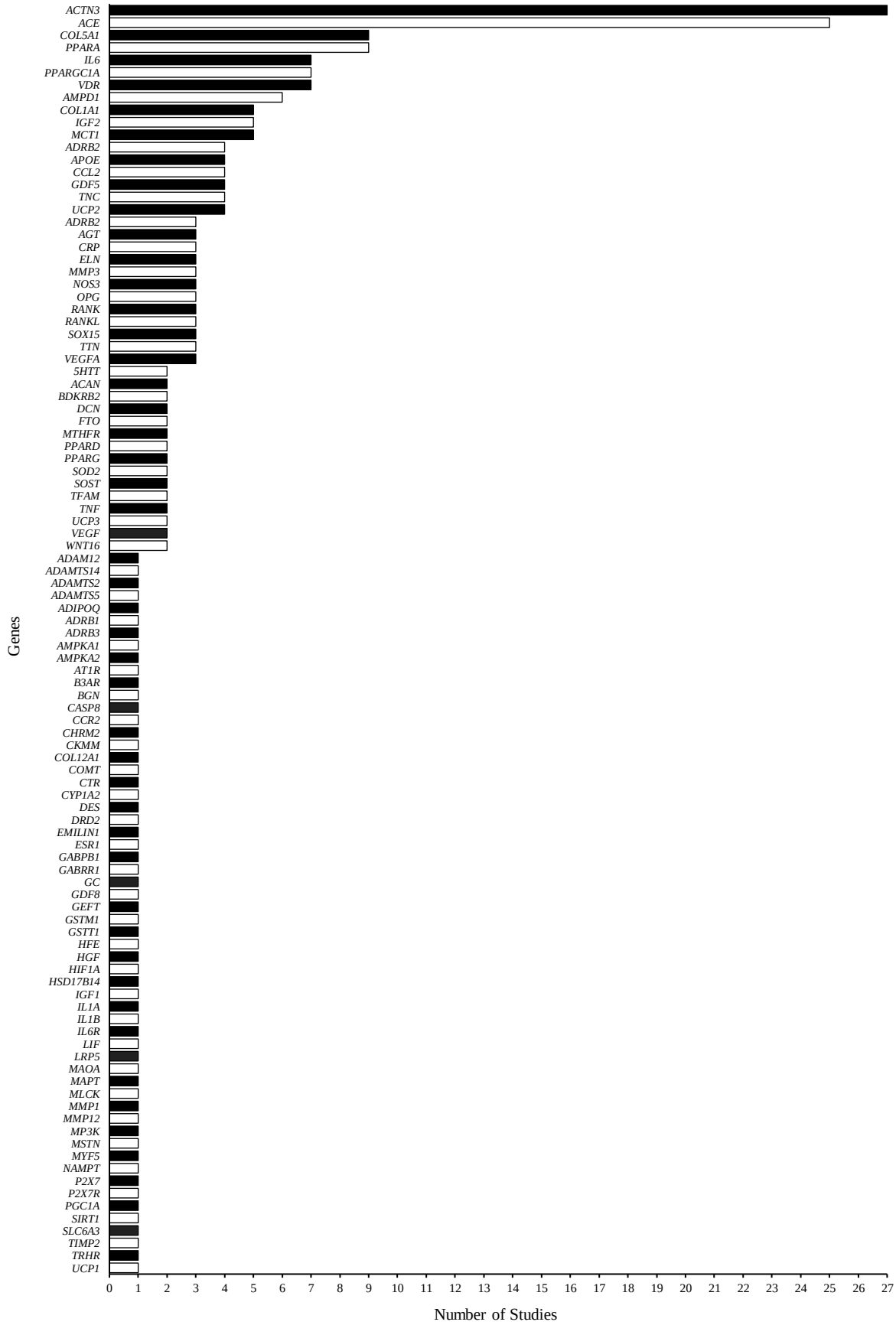


Figure 2.2. Names and frequencies of investigated genes.

2.4 Discussion

The aim of this review was to identify and evaluate the genetic association studies involving football players, whilst also assessing the current progress and methodological rigor of the research. This will provide researchers with a comprehensive evaluation of current methodological protocols and recommendations for future practice within the discipline. To the authors knowledge, this is the first review of methodological rigor in genetic association studies involving football players, thus providing a timely insight into an expanding area of research.

2.4.1 Progress

From the genetic research identified in this review, it seems that genetic association research in football is an ever-growing area of investigation. Furthermore, the types of studies and genes investigated are also expanding. The earliest investigation completed by Fatini et al. (2000), studied the association of the *ACE* (rs4646994) and angiotensin II receptor type 1 (*AT1R*, rs5186) genes with left ventricular hypertrophy. From the year 2000-2005, only one other study was conducted with footballers (Rizzo et al., 2003); who also studied the association of the *ACE* (rs4646994) gene with left ventricular hypertrophy. From the period of 2006-2010, seven new studies were conducted; although these studies chose to focus on new associations. The predominant research area during this time period changed to genetic associations with athletic status (Juffer et al., 2009; Oh et al., 2007; Santiago et al., 2008). However, new fields also included bone-phenotypes (Diogenes et al., 2010), physiological-phenotypes (Ginevičienė et al., 2010), and injury (Terrell et al., 210; Tierney et al., 2010). This was also accompanied by a growth in the number of novel genes studied, which progressed to ten. The next five years (2011-2015) saw the first explosive growth in genetic research within football. Twenty-six new studies were conducted, which included the first studies on epigenetics (Alfieri et al., 2015; Jeong et al., 2015) and the first psychological study (Filonzi et al., 2015). Furthermore, 28

additional novel genes were analysed during this time. Finally, from 2016 until the time of writing, 44 more studies have been conducted. This included another research focus, career progression (Coelho et al., 2018) and the inclusion of 65 additional novel genes. The current research and evidence presented here, of the growing interest in football genetics, corresponds with that of genetics and overall athletic performance research. For example, Ahmetov et al. (2016) showcased that from 1998-2015 a total of 155 SNP's had been associated with sporting physical performance. Of the identified polymorphisms, only 33% were identified in the first 12 years, with the remaining 77% identified in the last six years. This closely resembles the research presented here, as 62% of the genes studied in footballers have been within the last four years; as opposed to only 38%, in the preceding 16 years (see **Figure 2.3**). Therefore, it would appear research within football and genetics is growing at a substantial rate and may continue to increase.

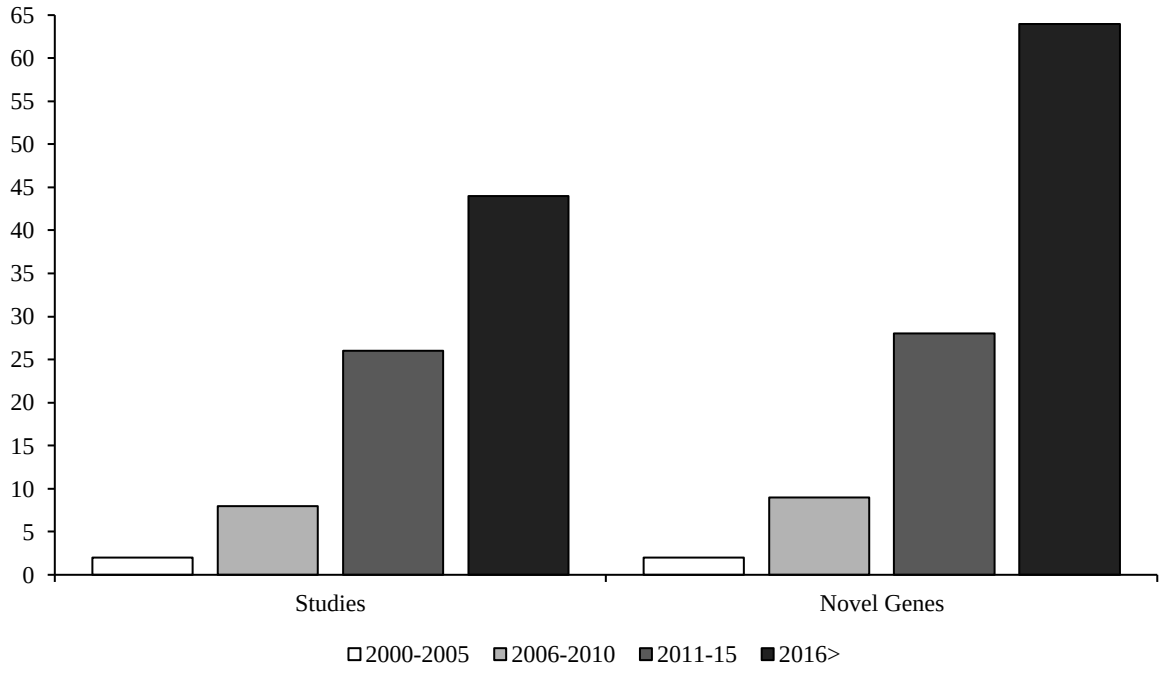


Figure 2.3. Increase in studies and novel genes in football genomics.

2.4.2 Methodological Rigor

Genetic association researchers should consider following the recommendations of the STREGA statement in the future. Only three studies (Cocci et al., 2019; Lulińska-Kuklik et al., 2018, 2019) in this review stated they attempted to adhere to STREGA recommendations. This resulted in inconsistent study designs and scores that ranged from 59-100% adherence. The main areas of weakness included; laboratory name and location omission, weak or omission of a limitation section, inclusion criteria vague or not included, bias (blinding) not mentioned, ethnicity and confounding factors not considered, Hardy-Weinberg equilibrium not mentioned, and information about funding not given. However, it must be noted that the STREGA recommendations were only published in 2009. Therefore, some studies were conducted before they could consider the recommendations. Furthermore, some elements of the STREGA recommendations are subjective and consequently scores could change depending on the assessors. Subsequently, STREGA recommendations are only meant to be utilised as a guideline to establish consistency of study-design. However, consistency will allow comparisons between studies to be evaluated easier and more accurately and therefore, future adherence to STREGA recommendations may help prevent the current limitations revealed in this review.

2.4.3 Research limitations

Analysis of the participant characteristics in this review reveals there are several methodological weaknesses across all study types within genetic association research in football; with the largest concern being the number of unknown participant characteristics. Three studies failed to report sex and four failed to report the performance level of their participants; whilst five multi-sport studies failed to report the number of football participants. Additionally, 13 studies failed to report age, whilst 22 studies failed to report the ethnicity of their participants. Consequently, 1,562 participants included in these studies have an unknown

ethnicity, which severely impacts the conclusions and ability to generalise across different ethnic populations. For example, Massidda et al. (2018) recently demonstrated that in a cohort of Italian (n = 266), Polish (n = 212), Lithuanian (n = 167), and Ukrainian (n = 49) footballers, Polish footballers had a distinct difference in frequency of the *MCT1* A1470T (rs1049434) polymorphism compared to the other ethnic groups. Therefore, for research to overcome the confounding factors of ethnic backgrounds, ethnicity must be consistently reported.

Aside from some of the unknown factors mentioned above, additional gaps in current research include a limited number of female, GWAS, and TGS studies. Female-only studies are valuable as a genetic polymorphism's allele frequency can be sex dependent (Ahmetov et al., 2013; Yang et al., 2003). Thus, while current research has elaborated on genetic associations with male footballers, little is known regarding females. Furthermore, whilst GWAS studies are limited due to the associated financial cost, they provide a hypothesis-free approach which could dramatically increase the number of novel genes identified in footballers (Visscher et al., 2012). Moreover, TGS studies consider the combined effect of genes, which is important as particular genes can have an impact on the expression of others (Williams & Folland, 2008). Furthermore, footballing performance is most likely the result of multiple genes and SNPs interacting with each other (gene-gene interactions), alongside the influence of environmental factors on those genes (gene-environment interactions) (Egorova et al., 2014). Additionally, it should be noted that 29 studies have included under 50 football players in their samples. Moreover, 32 studies were conducted on non-elite players and only 16 on youth players. Therefore, significantly higher sample sizes, senior athletes of a higher competitive level, and youth studies are needed in the future, in order to enhance practical implications (Eynon et al., 2011).

A lack of information concerning the on-field positions of players, is another significant limitation. Only five of the studies included the positional information of their football players.

Insufficient sample sizes may be the cause of this; or a lack of knowledge concerning inter-positional physiological demands. As previously noted, different positions in football have varying physiological requirements. For example, RSA is a key physiological factor within football and is correlated with team success (Oliver et al., 2009); however, RSA is more imperative to forward players (Di Salvo et al., 2007). An analysis of elite footballers in Europe showcased that forward players complete nearly double the number of sprints per game as central defenders and midfielders (Massidda et al., 2018). Furthermore, forwards covered twice the sprint distance of central midfielders. These findings were reinforced by Di Salvo et al. (2007), who discovered that after 20 La Liga and 10 Champions League matches, central defenders and midfielders perform far fewer sprints. Additionally, Aziz et al. (2008) discovered that forward players have a substantially higher RSA than midfielders and defenders. Therefore, future studies should at least display the number of players from each position in their studies. Furthermore, if studies have a sufficient sample size, they should separate their sample into separate categories based on their specific positions during a game. For example, Massidda et al. (2018) analysed players based on their on-field position and discovered that the A/A genotype of the A1470T (rs1049434) polymorphism in the solute carrier family 16 member 1 (*MCT1*) gene was twice as prominent in forward players than untrained subjects. This finding was most likely because the *MCT1* A allele is associated with higher lactate clearance (Cupeiro et al., 2016), which aids in sprinting performance. Therefore, as forward players sprint more and further during a game, the *MCT1* A allele could be advantageous for players in forward positions. Hence, not distinguishing between positions may inhibit the potential discovery of significant differences between players and controls; as physiological characteristics differ between positions, which indicates each position may be associated with distinct SNP alleles. Therefore, the number of players in a sample relative to their on-field

positions may significantly impact possible SNP associations and comparisons. However, these findings need to be replicated in other cohorts and ethnicities in order to be substantiated.

2.4.4 Future reviews

Of the genes that have been studied, *ACTN3* is the most represented, appearing in 27 of the included studies. This is followed closely by the *ACE* gene, which is involved in 25 of the included studies. *ACTN3* has also been conducted in the most fields of research including; athletic status, physiological phenotypes, injury, health, and career progression. The *ACE* gene has also appeared in; health, physiological phenotype, athletic status, and injury related studies. It is beyond the scope of this review to critically evaluate *ACTN3* and *ACE*'s impact on specific traits and overall performance within football players. Therefore, future reviews should aim to address this, given the contrasting results of the studies included in this review. For example, the R allele of the R577X polymorphism in the *ACTN3* gene has been positively (Galeandro et al., 2017; Santiago et al., 2008) and negatively (Massidda et al., 2014, 2015) associated with athletic status, along with a decreased incidence and severity of muscular injuries (Massidda et al., 2019). However, the contrasting results could be explained by a distinction in numerous variables (e.g., sample sizes, ethnicity, and sex). Therefore, independent reviews and meta-analyses could help inform researchers of the mechanisms responsible for these variations.

Furthermore, the most frequently studied fields of research (i.e., athletic status, physiological phenotypes, and injury) could be reviewed, given there are 20 or more studies in each. Although, it would be beyond the scope of this review, it would be useful to have an in-depth analysis of the variations in study designs and methodological considerations of those areas. For example, in the 20 studies that focused on athletic status, only three polymorphisms were replicated at least once: *ACE* I/D (n = 6); *ACTN3* R577X (n = 6); and, peroxisome proliferator activated receptor alpha (*PPARA*) intron 7 G/C (rs4253778; n = 5). However, whilst the D allele of the *ACE* I/D polymorphism was consistently associated with athletic

status all six times, discrepancies were revealed in the *ACTN3* and *PPARA* variants. Specifically, both alleles of the *ACTN3* R577X (R/X) and *PPARA* intron 7 (G/C) polymorphisms were associated with athlete status (R = 5, X = 1; G = 2, C = 3). As such, the lack of consistent replication between studies may be explained by the aforementioned methodological limitations. However, until a specific review is completed, the exact reason for these inconsistencies will not be revealed.

2.4.5 Future supplementary studies

The traits required to be a successful footballer are known to be multifactorial and involve both physiological and psychological attributes. So far, only three studies have investigated the potential role of psychological factors and genes, which have involved football players. Filonzi et al. (2015) was the first to do this through investigation of four genes (myostatin [*MSTN*] rs1805086; solute carrier family 6 member 4 [*5HTT*] rs25531; solute carrier family 6 member 3 [*SLC6A3*] VNTR; monoamine oxidase A [*MAOA*] rs6323), associated with psychological effects and elite performance. The authors reported that the *SLC6A3* gene is associated with sport success; due to the association of the *SLC6A3* gene with emotional control and anxiety management. Specifically, the authors found that the 9/9 genotype of the *SLC6A3* VNTR polymorphism was displayed five times as frequently in elite athletes compared to controls. Petito et al. (2016) then studied the *5HTT* gene in isolation, finding an association between the s/s genotype of the 5HTTLPR polymorphism and neuroticism, depression, and anxiety in elite athletes. Most recently, Cochrane et al. (2018) studied the association of three genes (apolipoprotein E [*APOE*] rs429358 + rs7412; catechol-O-methyltransferase [*COMT*] rs4680; dopamine receptor D2 [*DRD2*] rs1800497) on cognitive ability and concussion susceptibility in college athletes. The authors discovered that the *APOE* gene, e4 allele, was associated with significantly slower reaction times and the *COMT* gene, Val allele, was associated with significantly worse impulse control scores. There are some significant findings from these

studies, however in the context of footballers, there are some limitations. All studies were conducted with athletes from a range of different sports, with none focusing on footballers in isolation. Therefore, participants requiring distinct physiological and psychological attributes were inappropriately combined for analysis. Furthermore, Filonzi et al. (2015) only included four football players in their study, limiting their findings regarding footballers, whilst Petito et al. (2016) and Cochrane et al. (2018) failed to provide any information on the ethnicity of their participants. However, despite the limitations, these studies have provided valuable information for future studies, including identifying the potential genes to analyse in isolation or as part of a TGS. Therefore, future studies should attempt to replicate these findings on footballers and ensure adequate information is provided on participant characteristics.

In addition, Coelho et al. (2018) was the first study within football to investigate the relationship between genes and career progression. The authors did this by examining variations of the R577X polymorphism in the *ACTN3* gene between under-14, 15, 17, 20, and professional players. The authors reported there may be a trend towards a higher R/X genotype frequency in professional players than junior players, suggesting there could be a relationship between the R/X genotype and a football player's career progression. Specifically, the authors discovered there was an increased frequency of the R/X genotype in professional players compared to under-14 players. Therefore, the R/X genotype may be associated with performance attributes which lead to successfully obtaining a professional contract. However, although this study involved a sample of 353 male players, there are several limitations. The participants varied in ethnicity and the on-field positions of the players involved weren't disclosed. Ethnicity is particularly important regarding the R577X polymorphism, as a X/X genotype can facilitate a deficiency of a-actinin-3 protein (Yang et al., 2003), although this deficiency alters depending on the ethnic origin of the person. For example, the X/X genotype is prevalent in approximately 11% of Ethiopians compared to 18% in European ethnicities

(Yang et al., 2003). Therefore, populations of varied ethnic origins are already predisposed to different genotype frequencies which may influence results. Thus, any similar studies conducted in the future should aim to do so with an isolated ethnicity and analyse players according to their on-field position. However, it should also be noted that the population sample included very young players. As such, on-field positions may not be relevant, as during development a player may sample a variety of positions. Therefore, many players in the sample may not have had a fixed position. In addition, novel genes should be analysed either in isolation or as part of a TGS. This will facilitate an improved understanding of the specific mechanisms underpinning career progression in footballers.

2.4.6 Future novel empirical studies

While there is variation in the specific areas of research within genetic association research and football, there are some notable omissions and possible new areas for future research. There has yet to be a study into skill acquisition and ability involving football players. However, this has been investigated in other sports. For example, Jacob et al. (2018) investigated the association of nine SNPs on overall athleticism and skill in Australian Rules football. The authors revealed that the *ACE* (rs4343) gene was associated with overall athleticism and skill, while the brain derived neurotrophic factor (*BDNF*; rs6265), *DRD2* (rs1076560), and *COMT* (rs4680) genes were associated with kicking skill assessments. It is interesting to note that the *BDNF* gene is yet to be studied within football, despite having already been established as a main contributor in motor control and learning (Fritsch et al., 2010; Morin-Moncet., 2014). Therefore, it would be interesting if this study was replicated, to an extent, within football to validate these results. Furthermore, in a separate study, Jacob et al. (2019) also assessed the association of nine SNPs on match performance within Australian Rules football. This study revealed that the *BDNF* (rs6265) and *COMT* (rs4680) genes were associated with a player's direct game involvements, suggesting match performance and motor learning may have a

potential relationship. The study also highlighted the potential of the *BDNF* gene again within a team-sport, suggesting this study design would also be an interesting and novel addition to genetic research within football.

2.4.7 Advanced genomic technology and consortia

This review has emphasised the importance of utilising more advanced genomic technology (i.e., GWAS) and polygenic profiling (i.e., TGS) in the future. However, to sufficiently enhance our understanding of the molecular processes involved in football performance, the use of technology facilitating the identification of rare variants (i.e., whole-genome sequencing) is essential (Tanaka et al., 2016). For example, even a combination of polymorphisms may only explain a small proportion (1-3%) of phenotypic variability (Bouchard, 2015). Moreover, in order to fully understand all of the biological mechanisms underlying football performance, we will not only require genomic data, but also epigenomic data; potentially linked with transcriptomic and proteomic data (i.e., the application of 'omics') (Tanaka et al., 2016). However, accompanying a substantial increase in data is the number of multiple comparisons required during statistical testing. As such, this may result in an even greater genome-wide significance threshold, possibly 5×10^{-9} (Lin, 2019). Consequently, the sample sizes required to detect any potential association will need to be extensively larger and homogenous (Bouchard, 2015); which is already a significant issue limiting the field. A potential solution to this issue is the creation of an international football consortium (in accordance with similar suggestions of other researchers (Eynon et al., 2011, Mattsson et al., 2016). More specifically, the Athlome (GENESIS) Consortium (Pitsiladis et al., 2016), which contained a particularly applicable RugbyGene Project (Heffernan et al., 2015), may offer a suitable framework. In light of this, the authors have already begun collaborating with multiple football clubs and organisations, to generate a large genotypic and phenotypic database for the impending "Football Gene Project". The purpose of this project is to conduct a multi-

disciplinary investigation, utilising a range of genomic approaches, to identify novel genotype/phenotype associations in football, enhance our understanding of the biological mechanisms underpinning football performance, and ultimately facilitating greater individualised athlete development. As such, the authors welcome any interest from parties wishing to collaborate and enhance this database further in pursuit of this goal.

2.4.8 Limitations of this review

This review has several limitations which may have influenced the number of studies identified and included. Firstly, this review included papers which were only published in the English language. Therefore, studies published in other languages have been excluded. Furthermore, only the Pubmed and SPORTDiscus databases were searched for relevant studies. Therefore, studies published in other databases may have been missed. Moreover, as this study only included published papers, publication bias is also a limiting factor.

2.5 Conclusion

Overall, this review assessed the current progress and methodological rigor of genetic association research within football. Studies have increased at a substantial rate and significant progress has been made with regards to study type and genes investigated. However, research has predominately focused on the association of the *ACTN3* or *ACE* gene with performance parameters. Additionally, several sub-disciplines have also been explored; while some disciplines already require an independent review (i.e., injury), others require supplementary empirical research (i.e., psychological). Furthermore, there are several novel fields of research yet to be investigated in footballers which have been explored in other team-sports and revealed promising results (i.e., skill acquisition). There are also several methodological limitations within genetic association research in football which are currently preventing conclusive evaluations (i.e., population stratification). Therefore, adherence to STREGA guidelines, improved and consistent reporting, and an awareness of the influence of ethnicity and on-field

positions is advised. In addition, due to the relatively small sample sizes and the predominant use of CGASs, larger sample sizes and the utilisation of genome-wide and polygenic profiling approaches are recommended. The authors also wish to note that they realise the difficulty associated with some of the limitations raised in this review (i.e., sample size, ethnic conformation, and GWAS). Therefore, researchers should view the recommendations in this review as a guide towards future best practice, when more favourable conditions are available. The authors also introduced the Football Gene Project, which seeks to address several of these limitations through the collaboration of multiple professional footballing organisations, with the ultimate aim of facilitating greater individualised athlete development. The authors also suggest, in-line with the approach adopted by the aforementioned Athlome Project, that an international football consortium should be created. As such, this will enable the sharing of data between researchers to facilitate superior statistical power in future genetic association research in football.

3 The association of the *ACTN3* R577X and *ACE* I/D polymorphisms with athlete status in football: a systematic review and meta-analysis

McAuley, A. B. T., Hughes, D. C., Tsaprouni, L. G., Varley, I., Suraci, B., Roos, T. R., Herbert, A. J., & Kelly, A. L. (2021b). The association of the ACTN3 R577X and ACE I/D polymorphisms with athlete status in football: A systematic review and meta-analysis. *Journal of Sports Sciences*, *39*(2), 200–211. https://doi.org/10.1080/02640414.2020.1812195

Abstract

The aim of this review was to assess the association of *ACTN3* R577X and *ACE* I/D polymorphisms with athlete status in football and determine which allele and/or genotypes are most likely to influence this phenotype via a meta-analysis. A comprehensive search identified 17 *ACTN3* and 19 *ACE* studies. Significant associations were shown between presence of the *ACTN3* R allele and professional footballer status (OR = 1.35, 95% CI: 1.18-1.53) and the *ACE* D allele and youth footballers (OR = 1.18, 95% CI: 1.01-1.38) compared to a control group. More specifically, the *ACTN3* R/R genotype (OR = 1.48, 95% CI: 1.23-1.77) and *ACE* D/D genotype (OR = 1.29, 95% CI: 1.02-1.63) exhibited the strongest associations, respectively. These findings may be explained by the association of the *ACTN3* R/R genotype and *ACE* D/D genotype with power-orientated phenotypes and the relative contribution of power-orientated phenotypes to success in football. As such, the results of this review provide further evidence that individual genetic variation may contribute towards athlete status and can differentiate athletes of different competitive playing statuses in a homogenous team-sport cohort. Moreover, the *ACTN3* R577X and *ACE* I/D polymorphisms are likely (albeit relatively minor) contributing factors that influence athlete status in football.

3.1 Introduction

Actinin alpha 3 (ACTN3), a member of the actin family, is a sarcomeric protein which is greatly expressed in muscle tissue (Beggs et al., 1992). A function of the ACTN3 protein involves crosslinking fast-twitch (type II) actin filaments in skeletal muscle fibres (Mills et al., 2001). Thus, the expression of the ACTN3 protein in glycolytic skeletal muscle is thought to be a contributing factor to the generation of powerful and explosive muscle contractions; through optimal coordination of type II muscle fibres (Garton et al., 2014). The coding of the ACTN3 protein is controlled by the *ACTN3* gene, located on chromosome 11q13.2. A common genetic variant in the *ACTN3* gene has been identified which significantly alters the production of the ACTN3 protein (North et al., 1999). The genetic variation is a nonsense single nucleotide polymorphism (SNP) which can introduce a premature stop codon within the gene at position 577 (rs1815739) (Yang et al., 2003). Cytosine is the most common nucleotide at this position (i.e., **C**GA), which encodes the amino acid, arginine (R) (Mills et al., 2001). Alternatively, thymine can be possessed by an individual (i.e., **T**GA), producing the stop codon (X); resulting in X homozygotes being deficient in ACTN3 (North et al., 1999). As the *ACTN3* gene portrays a role in force production, it has been hypothesised that the performance of activities requiring extensive force production (i.e., sprinting, jumping, weightlifting) would be influenced by whether an individual possesses the R allele or R/R genotype. Many studies have reported that either the R/R genotype was over-represented, or the X/X genotype was underrepresented, in power-related sports (e.g., 100m sprint, rowing, speed skating, artistic gymnastics, sprint swimming, Olympic weightlifting) across American, Polish, Finnish, Italian, Japanese, Israeli, and Russian cohorts (Ahmetov et al., 2011; Chiu et al., 2011; Cięszczyk et al., 2011; Druzhevskaya et al., 2008; Eynon et al., 2009; Massidda et al., 2009; Mikami et al., 2014; Niemi & Majamaa, 2005; Papadimitriou et al., 2008; Roth et al., 2008). Indeed, Ma and colleagues (2013) conducted a meta-analysis on 23 studies involving power and endurance

athletes and discovered that the R allele was only associated with power athletes. Moreover, a recent meta-analysis on solely power athletes reported similar associations between the R allele and power athletes across 38 studies (Tharabenjasin et al., 2019).

Another commonly investigated gene in sport performance is the angiotensin I converting enzyme (*ACE*) gene. The angiotensin I converting enzyme catalyses the degradation of the inactive decapeptide angiotensin I, and subsequently generates the physiologically active peptide, angiotensin II; an oligopeptide of eight amino acids that binds to specific receptors in the body affecting several systems (Dzau, 1988; Munzenmaier & Greene, 1996). Angiotensin II can constrict blood vessels and stimulate aldosterone production, resulting in increased blood pressure, thirst, or the desire for salt. As such, the ACE enzyme is the most crucial component of the renin-angiotensin system (RAS), as it is a potent vasopressor and aldosterone-stimulating peptide which regulates blood pressure and fluid-electrolyte balance (Erdös & Skidgel, 1987). A polymorphism has been identified within intron 16 of the *ACE* gene, located on chromosome 17q23.3 (NC_000017.11), which results in a substantial variation of RAS activity (Danser et al., 1995; Rigat et al., 1990). The polymorphism is known as an insertion/deletion (indel) polymorphism, with the insertion (I allele) and deletion (D allele) representing the presence and absence of a 287-bp Alu-sequence, respectively. Specifically, the I allele has been associated with lower serum and tissue ACE activity, alongside an increased percentage of slow-twitch (type I) muscle fibres; whilst the D allele has been associated with higher circulating and tissue ACE activity, alongside greater strength and muscle volume and an increased percentage of type II muscle fibres (Danser et al., 1995; Rigat et al., 1990; Zhang et al., 2003). In the context of sport, the I allele has been frequently associated with elite endurance performance. Specifically, higher I allele frequencies have been reported in middle- and long-distance rowers, swimmers, road-cyclists, runners, mountaineers, cross-country skiers, and tri-athletes across a range of diverse cohorts (e.g., British, Australian, Croatian,

Russian, Spanish, Italian, Turkish, Polish, Japanese, Indian) (Alvarez et al., 2000; Cieszczyk et al., 2009; Gayagay et al., 1998; Min et al., 2009; Montgomery et al., 1998; Myerson et al., 1999; Nazarov et al., 2001; Scanavini et al., 2002; Shenoy et al., 2010; Turgut et al., 2004). Indeed, during the meta-analysis of Ma and colleagues (2013), the authors also assessed the influence of the *ACE* I/D polymorphism on endurance athletes over 25 studies, reporting that the I/I genotype was significantly associated with endurance athletes. However, the authors found no association between the *ACE* I/D polymorphism and power athletes. This may have been due to the large heterogeneity observed between studies (I^2 = >75%), most likely a result of not analysing the power athletes independently based upon ethnicity. Indeed, a more recent meta-analysis did conduct an ethnic specific analysis of power athletes and reported significant associations with the *ACE* D allele (Weyerstraß et al., 2018).

The collective results on the associations of the *ACTN3* R577X and *ACE* I/D polymorphisms, with athletic performance, complicate the implications for sports which require both power and endurance related traits, such as football. Football is an intermittent sport which requires optimal utilisation of both the aerobic and anaerobic systems (Bangsbo et al., 2006; Buchheit et al., 2010; Dellal et al., 2011; Mallo et al., 2015); and as such, predicting whether a power or endurance-orientated allelic variant may be preferable is not straightforward. Therefore, genetic association studies began to investigate whether a power or endurance-orientated genotype was more important in football by analysing polymorphisms such as *ACTN3* R577X and *ACE* I/D (Galeandro et al., 2017; Santiago et al., 2008; Ulucan et al., 2015). Moreover, studies began to assess if there was a difference between football players of various competitive playing levels (i.e., elite, non-elite; professional [PRO], non-professional [NP]) and controls (CON), in order to determine if *ACTN3* R577X or *ACE* I/D were associated with athlete status (Coelho et al., 2018; Egorova et al., 2014; Honarpour et al., 2017). Currently, there is no general consensus on the importance of *ACTN3* or *ACE* in

footballers, with studies reporting positive, negative, and contrasting allelic associations (Coelho et al., 2018; Egorova et al., 2014; Galeandro et al., 2017; Honarpour et al., 2017; Santiago et al., 2008; Ulucan et al., 2015). This is most likely because each gene or genotype has a small contribution to sporting performance and is dependent on numerous inter-individual variations (e.g., ethnicity and competitive playing level) (Monnerat et al., 2019; Wang et al., 2013). As a result, studies require large homogenous sample sizes in order to have sufficient statistical power and demonstrate significant associations and replications (Sagoo et al., 2009; Wang et al., 2013). However, studies within football genomics have notoriously small sample sizes and many are heterogenic multi-sport studies; mainly due to the unique population and limited access available (McAuley, Hughes, et al., 2021a). Therefore, to overcome the limitation of sample sizes and heterogeneity, a meta-analysis can be used to pool the results of single homogenous studies together (Sagoo et al., 2009). As such, the aim of this study was to assess if the *ACTN3* R577X and/or *ACE* I/D polymorphism are associated with athlete status in football by conducting a systematic review and meta-analysis.

3.2 Methodology

3.2.1 Search strategy and eligibility criteria

In accordance to the Preferred Reporting Items for Systematic reviews and Meta-Analysis (PRISMA) guidelines (Moher et al., 2009), the following search strategy was implemented. A comprehensive search of the Pubmed, SPORTDiscus, and MEDLINE databases was conducted on March 3rd 2020. For *ACTN3* the following Boolean search was used: ((football) OR (soccer)) AND (actn3) OR (alpha-actinin-3) OR (actinin-alpha-3) OR (R577X) OR (rs1815739)). For *ACE* the following Boolean search was used: ((football) OR (soccer)) AND (ace) OR (angiotensin I converting enzyme) OR (rs1799752) OR (rs13447447) OR (rs4341) OR (rs4646994)). Additionally, Google Scholar was searched using word combinations of the aforementioned Boolean searches, with no year restriction placed on any search. Furthermore,

reference lists of the identified articles were searched for additional relevant studies. At the initial screening stage studies were included if they: (1) were primary cohort or case-control investigations; (2) presented *ACTN3* R577X or *ACE* I/D genotype frequencies of footballers in isolation; and (3) were published in the English language. Therefore, studies were excluded if they: (1) were reviews; (2) presented *ACTN3* R577X or *ACE* I/D genotype frequencies of footballers combined with other sports; and (3) were published in a language other than in English.

3.2.2 Data extraction and analysis

We extracted the following data from all studies: first author's name and year of publication; number of footballers and CON; nationality and ethnicity; sex; age range; competitive playing level; type of study (cohort or case-control); and distribution of genotype frequencies in footballers and CON. Extracted data was then analysed in the following order: (1) pooled genotype frequencies of case-control studies in isolation; (2) pooled genotype frequencies of case-control studies combined with cohort studies (with an ethnically matched independent CON population added); and (3) sub-group analysis of ethnicity, sex, and level of competition. Independent CONs were added to cohort studies from the 1000 genomes database (https://www.internationalgenome.org) or an independent non-included study. Each cohort study was assigned a CON which was not used by any other study in the analysis to bolster the number of unique individuals and decrease selection bias. All of the included CON were also subjected to risk of bias before being included.

3.2.3 Risk of bias

After initial primary inclusion, all studies were subjected to Hardy–Weinberg equilibrium (HWE) via chi-square (significance level $p < .05$); culminating in the removal of all studies violating this significance threshold, as deviations from HWE in a CON group can indicate potential genotyping errors, selection bias and stratification (Salanti et al., 2005). Furthermore,

as each study is tested for HWE, Benjamini-Hochberg false discovery rate (FDR) is used to correct p-values and these adjusted p-values are displayed in-text and in the tables below (Benjamini & Hochberg, 1995). Additionally, as this review only uses published studies, publication bias could be a limiting factor. Therefore, Egger's test (Egger et al., 1997) in combination with funnel plots (Lau et al., 2006) were used to identify if publication bias was present and potentially skewing results (significance level $p < .05$).

3.2.4 Statistical analysis

To measure the strength of the association between the *ACTN3* R577X and *ACE* I/D polymorphisms and football players, odds ratios (OR), 95% confidence intervals (CI), and forest plots were used via either a fixed-effects model or random-effects model to identify the individual and pooled effects of the studies. Determining which model was appropriate for each analysis was based on the level of heterogeneity revealed via the I^2 (<50% = fixed-effects model; >50% = random-effects model) and Cochran's Q (Whitehead & Whitehead, 1991) statistical test (significance level at $p < .05$). Four genetic models were used to assess genotype and allele differences between football players and CONs: (a) allele contrast; (b) recessive; (c) dominant; and, (d) over-dominant. Additionally, pair-wise comparisons were also conducted. Finally, a sensitivity analysis was conducted via a leave-one-out approach in order to test the robustness of results (Viechtbauer & Cheung, 2010). All analyses were conducted using the MetaGenyo online Statistical Analysis System software (http://bioinfo.genyo.es/metagenyo/) (Martorell-Marugan et al., 2017).

3.3 Results

3.3.1 Search process

The systematic search processes initially identified 588 studies (*ACTN3*, *n* = 290; *ACE*, *n* = 298). Following the removal of duplicates and the screening of titles and abstracts, full-text assessment commenced; culminating in 17 *ACTN3* and 19 *ACE* studies being judged as

adequately meeting the predetermined inclusion criteria and subsequently being included in the final analysis (see **Figure 3.1** for full systematic search process).

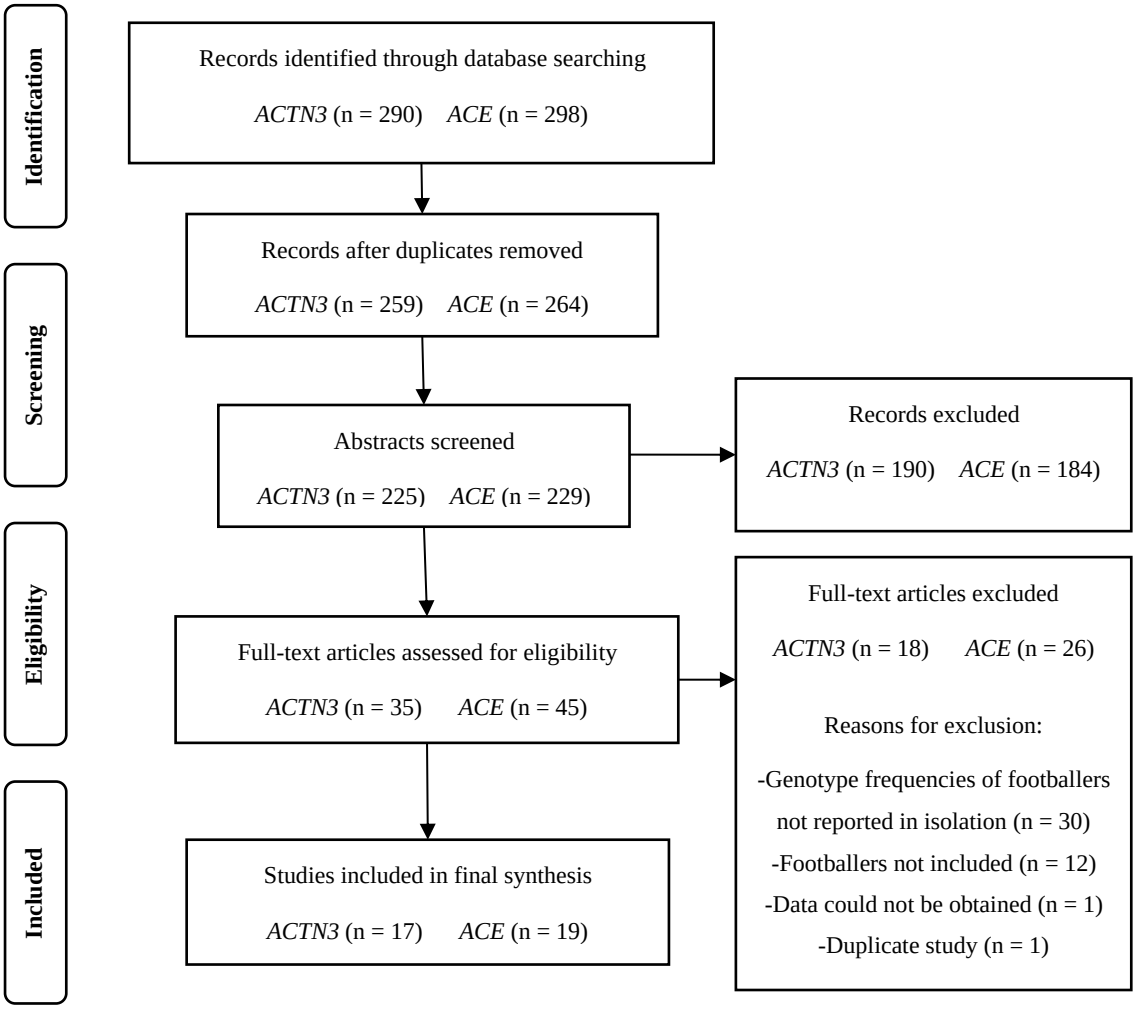


Figure 3.1. Flow diagram of systematic search process.

3.3.2 Study characteristics

The 17 *ACTN3* studies consisted of nine case-control and eight cohort studies respectively. There were a total of 1759 football players included across all studies (aged 10-37 years), with sample sizes ranging from 25-353. The most frequently studied nationality was Brazilian (*n* = 850), whilst the most frequently studied ethnicity was Caucasian (*n* = 587). Eleven studies included PRO players (*n* = 713; aged 17-37 years); four studies included NP players (*n* = 447; aged 14-30 years); and, two included both PRO and NP players (*n* = 599; aged 10-27 years).

The 19 *ACE* studies consisted of 15 case-control and four cohort studies. There were a total of 1925 football players included across all studies (aged 10-27 years), with sample sizes ranging from 25-353. The most frequently studied nationality was Brazilian (*n* = 709), whilst the most frequently studied ethnicity was Caucasian (*n* = 802). Ten studies included PRO players (*n* = 674; aged 17-26 years); seven studies included NP players (*n* = 652; aged 15-21 years); and, two included both PRO and NP players (*n* = 599; aged 10-27 years).

3.3.3 Risk of bias

Fifteen of the originally included 17 *ACTN3* studies remained after risk of bias assessment. The studies excluded, with reasons for exclusion, were as follows: (a) Massidda et al. (2014) failed HWE assessment, and (b) La Montagna et al. (2019) failed to provide the specific nationality and ethnicity of the footballers, thus an ethnically-matched CON could not be added. Seventeen of the originally included 19 *ACE* studies remained after risk of bias assessment. The following studies were excluded due to failing HWE assessment: Gineviciene et al. (2014) and Galeandro et al. (2017). In addition, the Bulgarian sub-cohort of Andreeva et al. (2014) was excluded from analysis due to also failing HWE assessment. In all genetic comparison models that identified a significant association with *ACTN3* and *ACE*, funnel plots did not reveal any signs of asymmetry and Egger's test did not detect a significant indication of publication bias.

3.3.4 Main analysis

Several significant associations were observed between the *ACTN3* R577X polymorphism and genetic comparison models with case-control studies in isolation: (a) allele contrast, (b) recessive, (c) dominant, (d) R/R vs. X/X, and (e) R/R vs. R/X. In addition, similar associations with the same genetic models were also observed in case-control studies combined with cohort studies; with the only exception being the additional significant association of R/X vs. X/X. The strongest association observed in both case-controls in isolation and combined with

cohorts was R/R vs. X/X (see **Table 3.1** for ORs and CIs of each *ACTN3* comparison). Sensitivity analysis assessed the robustness of the results and revealed that no study significantly altered pooled ORs. Conversely, no significant associations were observed between the *ACE* I/D polymorphism with case-controls in isolation or combined with cohorts using any genetic comparison model (see **Table 3.2** for ORs and CIs of each *ACE* comparison).

3.3.5 Sub analyses

The first sub-analysis assessed the independent associations of PRO players and NP players vs. controls (CON). Several significant associations were observed between the *ACTN3* R577X polymorphism and genetic comparison models in PROs: (a) allele contrast, (b) recessive, (c) dominant, (d) R/R vs. X/X, and (e) R/R vs. R/X. The strongest association observed was R/R vs. X/X. No significant associations were observed between the *ACTN3* R577X polymorphism and genetic comparison models in NPs vs. CON. Conversely, no significant associations were observed between the *ACE* I/D polymorphism and PRO players. However, several significant associations were observed between the *ACE* I/D polymorphism and NP players: (a) allele contrast, (b) dominant, and (c) I/D vs. D/D.

The second sub-analysis assessed the independent influence of ethnicity and nationality on associations. Due to the variance in geographical ancestry, Caucasian and Brazilian were the only ethnicity and nationality eligible for analysis. Several significant associations were observed between the *ACTN3* R577X polymorphism and genetic comparison models in Caucasians: (a) allele contrast, (b) recessive, (c) dominant, and (d) R/R vs. X/X.

Table 3.1. Statistical analysis of all studies investigating the *ACTN3* R577X polymorphism (*n* = 15).

Model		Association test				Heterogeneity		Bias
		OR	95% CI	*p*		*p*	I^2	*Egger's*
Case-Control								
Allele contrast		1.30	1.15-1.46	< **.001**		0.68	0.00	0.88
Recessive		1.42	1.20-1.68	< **.001**		0.22	0.24	0.87
Dominant		1.35	1.06-1.72	**.016**		0.33	0.13	0.97
Over-dominant		0.84	0.66-1.07	.160		0.05	0.45	0.93
RR vs. XX		1.67	1.28-2.17	< **.001**		0.63	0.00	0.98
RR vs. RX		1.38	1.07-1.78	**.013**		0.08	0.41	0.76
RX vs. XX		1.17	0.90-1.51	.230		0.11	0.38	0.92
Case-Control & Cohort								
Allele contrast		1.26	1.15-1.38	< **.001**		0.72	0.00	0.48
Recessive		1.31	1.15-1.50	< **.001**		0.23	0.18	0.89
Dominant		1.40	1.17-1.69	< **.001**		0.49	0.00	0.64
Over-dominant		0.93	0.78-1.12	.450		0.05	0.38	0.75
RR vs. XX		1.60	1.31-1.96	< **.001**		0.72	0.00	0.62
RR vs. RX		1.23	1.02-1.49	**.029**		0.08	0.34	0.83
RX vs. XX		1.29	1.06-1.57	**.010**		0.18	0.24	0.69
PRO vs. NP								
Allele contrast	PRO	1.35	1.18-1.53	< **.001**		0.77	0.00	0.34
	NP	1.07	0.90-1.27	.440		0.90	0.00	0.30
Recessive	PRO	1.48	1.23-1.77	< **.001**		0.30	0.14	0.70
	NP	1.00	0.78-1.29	.970		0.74	0.00	0.46
Dominant	PRO	1.40	1.09-1.81	**.010**		0.16	0.30	0.83
	NP	1.27	0.91-1.76	.160		1.00	0.00	0.34
Over-dominant	PRO	0.84	0.63-1.12	.240	**.020**		0.51	0.89
	NP	1.13	0.89-1.43	.330		0.79	0.00	0.66
RR vs. XX	PRO	1.77	1.34-2.34	< **.001**		0.57	0.00	0.61
	NP	1.22	0.84-1.76	.290		0.95	0.00	0.21
RR vs. RX	PRO	1.41	1.07-1.86	**.014**		0.07	0.40	0.98
	NP	0.94	0.72-1.23	.680		0.72	0.00	0.52
RX vs. XX	PRO	1.22	0.79-1.87	.360	**.028**		0.50	0.91
	NP	1.30	0.92-1.84	.140		0.99	0.00	1.00
Caucasian & Brazilian								
Allele contrast	Caucasian	1.32	1.14-1.53	< **.001**		0.39	0.05	0.73
	Brazilian	1.23	1.07-1.41	**.003**		0.45	0.00	0.47
Recessive	Caucasian	1.41	1.14-1.74	**.001**		0.11	0.42	0.67
	Brazilian	1.19	0.98-1.44	.070		0.64	0.00	0.69
Dominant	Caucasian	1.49	1.11-2.00	**.008**		0.79	0.00	0.82
	Brazilian	1.52	1.15-1.99	**.003**		0.17	0.35	0.92
Over-dominant	Caucasian	0.91	0.75-1.12	.380		0.11	0.42	0.69
	Brazilian	1.07	0.89-1.29	.480		0.59	0.00	0.45
RR vs. XX	Caucasian	1.71	1.24-2.36	**.001**		0.60	0.00	0.60
	Brazilian	1.61	1.19-2.16	**.002**		0.23	0.27	0.94
RR vs. RX	Caucasian	1.36	0.96-1.92	.080		0.08	0.47	0.61
	Brazilian	1.08	0.88-1.32	.460		0.66	0.00	0.85
RX vs. XX	Caucasian	1.34	0.98-1.83	.060		0.61	0.00	0.97
	Brazilian	1.48	1.10-1.98	**.009**		0.17	0.35	0.98

Note. Allele contrast (R vs. X); Recessive (R/R vs. R/X-X/X); Dominant (R/R-R/X vs. X/X); Over-dominant (R/X vs. R/R-X/X); OR = Odds Ratio; CI = Confidence Interval; PRO = Professional; NP = Non-Professional.

Table 3.2. Statistical analysis of all studies investigating the *ACE* I/D polymorphism (*n* = 17).

Model		Association test			Heterogeneity		Bias
		OR	95% CI	*p*	*p*	I^2	*Egger's*
Case-Control							
Allele contrast		0.88	0.73-1.07	.21	**.001**	0.61	0.16
Recessive		0.84	0.69-1.03	.09	0.44	0.01	0.72
Dominant		0.84	0.62-1.15	.28	< **.001**	0.66	0.09
Over-dominant		0.93	0.74-1.16	.51	**.040**	0.44	0.04
II vs. DD		0.82	0.58-1.17	.27	**.025**	0.47	0.19
II vs. ID		0.98	0.79-1.21	.82	0.65	0.00	0.27
ID vs. DD		0.84	0.62-1.14	.26	**.002**	0.60	0.08
Case-Control & Cohort							
Allele contrast		0.87	0.75-1.02	.09	**.004**	0.53	0.10
Recessive		0.86	0.72-1.03	.10	0.69	0.00	0.69
Dominant		0.82	0.63-1.07	.14	< **.001**	0.62	0.05
Over-dominant		0.90	0.73-1.10	.31	**.019**	0.45	0.03
II vs. DD		0.79	0.60-1.05	.10	0.08	0.34	0.12
II vs. ID		1.00	0.83-1.22	.97	0.75	0.00	0.21
ID vs. DD		0.81	0.62-1.06	.13	**.001**	0.58	0.05
PRO vs. NP							
Allele contrast	PRO	1.01	0.86-1.19	.880	0.96	0.00	0.67
	NP	0.85	0.73-0.99	**.044**	0.15	0.35	0.45
Recessive	PRO	0.97	0.72-1.29	.820	0.99	0.00	0.99
	NP	0.88	0.65-1.18	.390	0.20	0.29	0.38
Dominant	PRO	1.05	0.82-1.34	.710	0.86	0.00	0.73
	NP	0.78	0.61-0.98	**.032**	0.25	0.22	0.44
Over-dominant	PRO	1.07	0.85-1.34	.570	0.81	0.00	0.62
	NP	0.86	0.69-1.07	.180	0.22	0.27	0.95
II vs. DD	PRO	1.05	0.74-1.49	.770	0.99	0.00	0.95
	NP	0.77	0.55-1.08	.140	0.22	0.26	0.26
II vs. ID	PRO	0.94	0.69-1.27	.690	0.98	0.00	0.85
	NP	0.97	0.71-1.33	.850	0.20	0.28	0.51
ID vs. DD	PRO	1.05	0.80-1.36	.730	0.80	0.00	0.55
	NP	0.76	0.59-0.97	**.030**	0.38	0.07	0.56
Caucasian & Brazilian							
Allele contrast	Caucasian	0.81	0.57-1.14	.230	**.004**	0.74	0.04
	Brazilian	0.85	0.70-1.03	.090	0.37	0.07	0.63
Recessive	Caucasian	0.75	0.57-1.00	.050	0.61	0.00	0.50
	Brazilian	0.83	0.60-1.14	.250	0.25	0.26	0.07
Dominant	Caucasian	0.75	0.42-1.34	.340	< **.001**	0.82	0.03
	Brazilian	0.77	0.55-1.06	.110	0.17	0.38	0.73
Over-dominant	Caucasian	0.83	0.53-1.29	.400	**.009**	0.70	0.08
	Brazilian	1.05	0.64-1.71	.850	**.018**	0.67	0.29
II vs. DD	Caucasian	0.65	0.38-1.10	.110	0.08	0.52	0.08
	Brazilian	0.80	0.52-1.22	.290	0.31	0.17	0.58
II vs. ID	Caucasian	1.05	0.77-1.44	.750	0.59	0.00	0.28
	Brazilian	0.88	0.63-1.23	.450	0.10	0.48	0.19
ID vs. DD	Caucasian	0.76	0.41-1.42	.390	< **.001**	0.82	0.07
	Brazilian	0.82	0.47-1.42	.480	0.09	0.51	0.57

Note. Allele contrast (I vs. D); Recessive (I/I vs. I/D-D/D); Dominant (I/I-I/D vs. D/D); Over-dominant (I/D vs. I/I-D/D); OR = Odds Ratio; CI = Confidence Interval; PRO = Professional; NP = Non-Professional.

Likewise, several significant associations were also observed between the *ACTN3* R577X polymorphism and genetic comparison models in Brazilians: (a) allele contrast, (b) dominant, (c) R/R vs. X/X, and (d) R/X vs. X/X. The strongest association observed in both Caucasians and Brazilians was R/R vs. X/X. Conversely, no significant associations were observed between the *ACE* I/D polymorphism and Caucasians or Brazilians.

3.4 Discussion

The aim of this study was to assess if the *ACTN3* R577X and/or *ACE* I/D polymorphism were associated with athlete status in football; and if so, determine which allele and/or genotypes are most likely to influence this phenotype via a meta-analysis. To the author's knowledge this is the first review to do this within football; and moreover, it is the first meta-analysis in a homogenous team-sport cohort.

3.4.1 Main analysis

Following meta-analysis, this review identified several associations between the *ACTN3* R577X polymorphism and all football players (PRO & NP combined) vs. CON. In summary, the main associations were observed in football players possessing the R allele, with the strongest association being the R/R vs. X/X genotype. These associations were observed in case-control studies in isolation and remained similar with the addition of cohort studies. Although, with the significant increase in sample size from cohort studies the CI was reduced, possibly indicating a more accurate estimation. The only notable difference between genetic comparison models was that the R/X vs. X/X pair-wise comparison was statistically significant with cohorts added vs. non-significant in case-controls in isolation. Therefore, only the over-dominant model remained non-significant in both case-control studies in isolation and with the addition of cohort studies; most likely due to the strength of the association of the R/R genotype. As such, the results of this analysis showcase that in football players, similar to power athletes, there is an overrepresentation of the *ACTN3* R/R genotype. For instance, this

can be illustrated by the similar recessive model and dominant model findings of the present study and Ma and colleagues (2013), respectively. This can likely be explained by the combination of several factors: (a) the number and frequency of powerful actions performed in a game (i.e., 1000-1400 acyclical bursts of activity, including; jumps, tackles, shots at goal, changes of direction, and, sprints), with a high-intensity sprint occurring every ~70s (Bradley et al., 2009; Stølen et al., 2005), (b) the contribution of the ability to repeat higher intensity actions to success in football (i.e., league position, goals scored, goals prevented, duel success) (Little & Williams, 2005; Oliver et al., 2009), and (c) the association of the R/R genotype with power-orientated phenotypes (i.e., vertical jumping and 10-30m sprints) (Dionísio et al., 2017; Pimenta et al., 2013). However, it is important to recognise that there are many other genetic polymorphisms, with each likely influencing performance to a limited extent (Monnerat et al., 2019; Wang et al., 2013). Therefore, football players, power athletes, and indeed other team-sports, may possess contrasting allele and genotype frequencies in other polymorphisms. Hence, the results of this review only indicate a difference between football players and non-athletic CONs. However, this difference appears to be greatly mediated by competitive playing level, which may have contributed to the heterogeneity present between *ACE* studies; and as a result, the non-significant associations between *ACE* and PRO & NP players combined. Indeed, the results of the subsequent sub-analysis on competitive playing level revealed that *ACE* may play an important role in determining athlete status earlier in athlete development. Therefore, although *ACE* was not associated with athlete status in the main analysis, it is still an important genetic variant in football, depending on specific competitive playing levels.

3.4.2 Sub-analyses

The importance of the competitive playing level of footballers in this review was assessed via separating PRO players and NP players. PRO players were classified as players that studies specifically described as playing at a professional level, whereas NP players were

classified as players playing at a semi-professional, amateur, or youth level. This particular method of categorisation was chosen, as opposed to elite vs. non-elite, to circumvent the problematic classification issue of 'eliteness' (i.e., what constitutes elite status and how do we define it?) (Kirkland & O'Sullivan, 2018; Swann et al., 2015). For example, players may be internationals who represent their country (normally classed as elite), however what if their country is near the bottom of the international rankings? Furthermore, players may play for a European club positioned 1st in their country's highest league (normally classed as elite), but what if the country and club both have a low UEFA coefficient ranking? Moreover, how do we define the highest performing youth players? Given that no irrefutable solution has been provided and no general consensus has been agreed, the term 'elite' is still inconsistently utilised in the literature to describe varying standards of performance (Kirkland & O'Sullivan, 2018; Swann et al., 2015). Therefore, to reduce between-study-heterogeneity, the authors chose to compare PROs and NPs (limitations of this approach are discussed below).

Firstly, the results of this analysis revealed that the *ACTN3* R577X polymorphism was associated with PRO players vs. CON. However, no statistically significant difference was observed between NP players and CON. Moreover, the strength of the association between the R allele, and specifically the R/R genotype, was even stronger in PRO players when separated from NP players. As such, these results indicate that the R allele is a likely (albeit small) contributing factor towards attaining PRO status in football. Secondly, whilst the *ACE* I/D polymorphism was not associated with PRO players vs. CON, it was associated with NP players vs. CON. To be specific, NP players were more likely to possess a D allele or D/D genotype. This is perhaps more easily demonstrated via inverse statistical analysis: (1) D vs. I (OR = 1.18, 95% CI: 1.01-1.38); (2) D/D vs. D/I-I/I (OR = 1.29, CI: 1.02-1.63); and, (3) D/D vs. D/I (OR = 1.32, CI: 1.03-1.69). Interestingly, the NP players consisted solely of youth players in the *ACE* I/D studies (n = 652; aged 15-21 years), showcasing that the D allele is only

overrepresented specifically in youth football. Furthermore, the differences between PROs and NPs may have increased further given that Egorova and colleagues (2014) was not included in this analysis, due to not displaying the individual genotype frequencies of PRO and NP players in their study. However, the authors did report that when separated by competitive playing level, only elite players displayed a significantly higher frequency of the *ACTN3* R allele compared to CON (81.1%, $p < .001$); whilst conversely, only youth players displayed a significantly higher frequency of the *ACE* D allele compared to CON (78.8%, $p < .001$). As such, the observed cumulative evidence suggests that whilst possessing the *ACE* D allele is potentially more beneficial to young players, the *ACTN3* R allele is the only allele which likely has a (minor) role in attaining PRO status (see **Figure 3.2**). However, it is important to note the use of terms such as "albeit small" and "minor" when interpreting these findings. Polymorphisms account for very little of the inter-individual variance in complex traits such as athlete status (Wang et al., 2013). Indeed, even a combination of 97 polymorphisms (at $p = 5\times10^{-8}$ or better) could only account for 2.7% of body mass index variance between individuals (Locke et al., 2015). Therefore, as evidenced by the relatively small odds ratios in this study, *ACTN3* and *ACE* similarly play a minor role in determining athlete status in football.

The explanation for the observed associations between the *ACE* D allele and *ACTN3* R allele with youth and PRO players respectively is challenging. Both alleles have been previously associated with strength/power orientated sports and general strength/power characteristics (i.e., increased percentage of type II muscle fibres, strength, and muscle mass) (Ahmetov et al., 2011; Zhang et al., 2003). More specifically, in football cohorts both alleles have been positively associated with greater countermovement and squat jump performance, and faster 10m, 20m, and 30m sprint times (Dionísio et al., 2017; Pimenta et al., 2013).

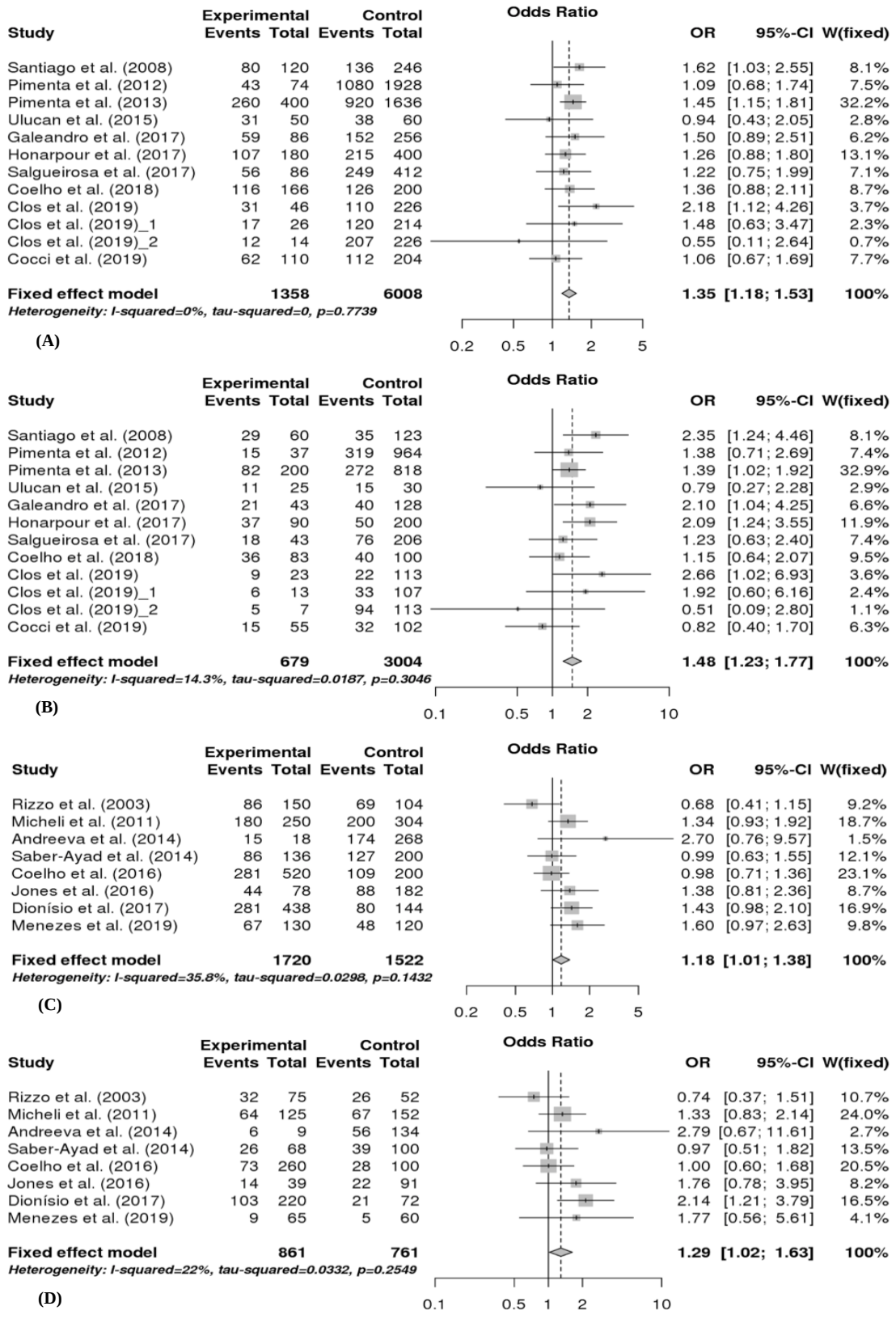


Figure 3.2. Forest plots of significant PRO and NP comparison models. (A) *ACTN3* R vs. X allele in PROs; (B) *ACTN3* R/R vs. R/X-X/X in PROs; (C) *ACE* D vs. I allele in NPs; (D) *ACE* D/D vs. D/I-I/I in NPs.

As such, both alleles appear to be associated with the same side of the endurance-power continuum in football. Therefore, it is interesting that each allele is independently associated with different competitive playing levels. However, perhaps the categorisation method employed to distinguish competitive playing levels in this review could possibly be responsible. For example, whilst heterogeneity between studies in the PRO vs. NP analysis was mostly small, the age of NPs in *ACTN3* ranged from 14-30 years; whereas the age of NP players in *ACE* ranged from 15-21 years. As such, it is possible that the NP *ACE* players had greater performance levels than the NP *ACTN3* players, as the NP *ACE* players may play at the highest youth level in their respective age groups, whilst the *ACTN3* players may not. Indeed, several previous studies have demonstrated that 'elite' youth players, aged 14-17 years, have outperformed non-elite players in acceleration, speed and jumping assessments (Gil et al., 2007; Waldron & Murphy, 2013). Although, it could be argued that grouping young players (14-17 years) as elite and non-elite is not appropriate as factors such as maturation status may influence performance more than *ACTN3* or *ACE* genotype. However, in truth, numerous potential explanations exist, including: (a) differences in competitive performance levels, (b) variations in geographical ancestry, (c) disparities in the distributions of players relative to their on-field position within samples, (d) distinct methods of genotyping, (e) individual gene-gene interactions, and (f) separate gene-environment interactions. Moreover, it is also important to note that the *ACE* gene is part of a very complex pathway, with serval interactions that can influence it's activity; whilst *ACTN3* represents the presence/absence of a structural protein (Dzau, 1988; Munzenmaier & Greene, 1996). As such, the specific cause of the distinction between the *ACTN3* R577X and *ACE* I/D polymorphisms with PRO and NP players requires further research to elucidate these findings.

In addition to competitive playing level, sub-analysis also assessed the influence of ethnicity and geographical ancestry on observed associations. Therefore, studies, and samples

within studies, were separated based upon their reported ethnic heritage and nationality. After separation, Caucasians and Brazilians collectively represented 81% and 79% of *ACTN3* and *ACE* studies respectively. Therefore, only these could be assessed via comparison analysis. Firstly, in relation to the *ACE* I/D polymorphism, no significant associations were observed in either Caucasians or Brazilians. However, in relation to the *ACTN3* R577X polymorphism, several significant associations were observed in both Caucasians and Brazilians between the R allele and football players vs. CON. However, in models with significant associations in both Caucasians and Brazilians (allele contrast, dominant, R/R vs. X/X), Caucasians displayed higher ORs. This suggests the R allele, and more significantly the R/R genotype, appear to be of greater importance to footballers of Caucasian heritage. This suggestion is bolstered by the significant associations observed solely between Caucasians and a recessive model, and between Brazilians and the pair-wise comparison of the R/X vs. X/X genotypes. This reveals that the R/X genotype may be more associated with footballing status in Brazilians, whilst the R/R genotype may be more associated with footballing status in Caucasians.

The possible cause of the disparities in ORs between the R allele in Brazilians and Caucasians may be related to the ethnic diversity of the Brazilian population; and more specifically, the proportion of individuals with African ancestry. Individuals of African ancestry have greater frequencies of the R/R genotype (~78%) and significantly lesser frequencies of the X/X genotype (~1%), irrespective of athlete status, compared to the estimated frequencies of the world population (R/R ~40%; X/X ~18%) (1000 Genomes Project Consortium et al., 2015; Scott et al., 2010; Yang et al., 2007). As such, the *ACTN3* R577X polymorphism does not differentiate elite athletes from CON of African ethnicity in either power or endurance-orientated sports (Scott et al., 2010; Yang et al., 2007). Using ancestry informative markers, previous studies have reported that the Brazilian population are formed of ~65% European SNPs, ~22% African SNPs, and ~13% Amerindian SNPs (Barbosa et al.,

2017). Furthermore, these proportions vary in each independent region of Brazil. For example, it is estimated that the population of Rio de Janeiro is formed of ~55% European SNPs, ~31% African SNPs, and ~14% Amerindian SNPs (Saloum de Neves Manta et al., 2013). As such, given the genetic similarity of the Brazilian population with the genetic profile of African ethnicities, this potentially explains why the R/R genotype may contribute to footballing status in Brazil to a lesser extent.

3.4.3 Strengths and limitations

This study's findings are reinforced due to the small-moderate heterogeneity identified between studies. Only two of the 28 total measurements regarding *ACTN3* displayed an I^2 value over 50% and violated Cochran's Q statistical threshold ($p < .05$); although neither of these two measurements displayed a statistically significant result. Likewise, neither of the three associations identified between the *ACE* I/D polymorphism and NP players displayed significant heterogeneity. Furthermore, no suggestion of publication bias was identified regarding the observed significant associations, through analysis of funnel plots or Egger's test ($p < .05$).

This review is not without limitations and each must be duly considered when interpreting the findings. Firstly, this review attempted to include both male and female football players, but unfortunately the authors could only locate one study involving female players (Andreeva et al., 2014) which met the pre-defined inclusion criteria. Therefore, the results from this review mainly apply to male football players. Secondly, the ethnic and geographic implications of this review mainly concern Brazilians and Caucasians, due to both representing the majority of the football players included in this review (*ACTN3* = 48% & 33% respectively; *ACE* = 37% & 42% respectively). Thus, research is still required on players of diverse geographical ancestry. Thirdly, the division of players into PRO and NP categories was chosen to circumvent the well-known issues surrounding what constitutes 'eliteness' in sport;

however, the PRO level still encompasses substantial disparities in standards of play (e.g., English Premier League vs. English Football League Two). As such, this makes it difficult to objectively quantify player ability and consequently associate the findings with specific levels of performance. Finally, the on-field positions of footballers were rarely reported in the included studies. This is important considering the previously reported inter-positional associations between the *ACTN3* R577X and *ACE* I/D polymorphism and football players, such as forwards and goalkeepers possessing significantly higher frequencies of the *ACTN3* R allele and *ACE* D allele, respectively (Egorova et al., 2014). Therefore, it is unknown whether the associations in this review were influenced by the number of players included in each study relative to their on-field positions.

3.4.4 Practical applications

Limited evidence exists supporting the practical application of genetic information with athlete development. For example, it has been reported that individuals possessing the *ACTN3* R/R and *ACE* D/D genotypes may have an improved adaptive response in strength and power with heavy resistance training (Delmonico et al., 2007; Erskine et al., 2014). As such, Kikuchi and Nakazato (2015) suggested that individuals possessing power-orientated genetic variants, such as *ACTN3* R/R, may benefit more from resistance training consisting of high weights with low-repetitions for strength and power. Jones and colleagues (2016) investigated this theory and reported that in 39 football players the *ACE* D/D and *ACTN3* R/R genotypes demonstrated a greater improvement in countermovement jump performance in response to high-weight low-repetition resistance training; whereas, the *ACE* I/I and *ACTN3* X/X genotypes demonstrated a greater improvement with low-load high-repetition resistance training. However, this study comprised a small sample size, which consequently increases the likelihood of type 1 error occurrence (Karanikolou et al., 2017). Furthermore, to the author's knowledge, the results presented by Jones and colleagues (2016) have yet to be replicated in an independent cohort.

Therefore, further intervention studies are required in order to establish a strong evidence base, before practical recommendations can be proposed regarding the results of this review. However, once a strong evidence base has been established which supports genetic-based programme design, this review can be used alongside a number of other extensively evidenced/researched genotypes to form the basis of genetic-based training. Indeed, studies have also showcased that differences in biomarkers of muscle damage, hormones, and inflammatory responses, may be influenced by genotype variation (Coelho et al., 2019; Pimenta et al., 2012). Therefore, the utilisation of genetic information in the future may not only aid in optimising training adaptation, but also the planning of athlete workloads and recovery.

3.5 Conclusion

The results of this review provide further evidence that individual genetic variation likely contributes towards athlete status. Our findings suggest that genetic variation can differentiate athletes of different competitive playing statuses in a homogenous team-sport cohort. Specifically, this review has showcased that the R allele and R/R genotype of the *ACTN3* R577X polymorphism are overrepresented in PRO football players; whereas the D allele and D/D genotype of the *ACE* I/D polymorphism are overrepresented in youth players. These overrepresentations may be explained by the association of the *ACTN3* R/R genotype and *ACE* D/D genotype with power-orientated phenotypes and the relative contribution of these power-orientated phenotypes to success in football. As such, the *ACTN3* R577X and *ACE* I/D polymorphisms are likely (albeit relatively minor) contributing factors which influence athlete status in football.

4 A systematic review of the genetic predisposition to injury in football

McAuley, A. B. T., Hughes, D. C., Tsaprouni, L. G., Varley, I., Suraci, B., Roos, T. R., Herbert, A. J., Jackson, D. T., & Kelly, A. L. (2022d). A systematic review of the genetic predisposition to injury in football. *Journal of Science in Sport and Exercise*. Advance online publication. https://doi.org/10.1007/s42978-022-00187-9

Abstract

Injury has a profound effect on a football team's potential success and economic burden. Considerable research in football has been dedicated to finding ways to decrease injuries, which has seen a rising interest in genetic susceptibility. The aim of this review was to synthesise genetic association studies investigating injury involving football players to identify which genetic variants have the most empirical evidence to date. A comprehensive search of the Pubmed, SPORTDiscus, and MEDLINE databases until March 11th 2022 identified 34 studies. There were 33 candidate gene studies and one genome-wide study, with 9642 participants across all studies. Ninety-nine polymorphisms were assessed within 63 genes. Forty-one polymorphisms were associated with injury once. Three polymorphisms had their specific allelic associations with injury replicated twice in independent cohorts: *ACTN3* (rs1815739) X/X genotype was associated with an increased susceptibility to non-contact muscle injuries, *ACAN* (rs1516797) G allele was associated with increased susceptibility to anterior cruciate ligament (ACL) injuries, and *VEGFA* (rs2010963) C/C genotype was associated with an increased susceptibility to ACL and ligament or tendon injuries. However, several methodological issues (e.g., small sample sizes, cohort heterogeneity, and population stratification) limit the reliability and external validity of findings. At present, the evidence base supporting the integration of genetic information as a prognostic or diagnosis tool for

injury risk in football is weak. Future participation of organisations in international consortia is suggested to combat the current methodological issues and subsequently improve clarity concerning the underlying genetic contribution to injury susceptibility.

4.1 Introduction

Based on a squad of 25 players, a professional football team will sustain approximately fifty time-lost injuries a season (Ekstrand et al., 2011). Consequently, each player will likely sustain two injuries per season that affect their availability for match selection. A recent meta-analysis in professional football reported an injury incidence rate of 8.1/1000 hours of exposure (López-Valenciano et al., 2020). This is problematic, as injuries which result in match-minutes lost have a strong association with team success. To be specific, player availability is strongly correlated with league position, total points, games won, and goals scored (Eirale et al., 2013; Hägglund et al., 2013). Injuries also significantly impact a club's finances. For example, a professional player injured for a month costs a club competing in the men's UEFA Champions League approximately €500,000 (Ekstrand, 2013). Moreover, across five seasons (2012/2013 to 2016/2017) in the English Premier League, it was estimated that injuries cost clubs an average of £45,000,000 per season (Eliakim et al., 2020). Injury risk is multifactorial, but recent research into the identification of risk factors that predispose injury has seen a rise in interest in genetic susceptibility (McAuley, Hughes, et al., 2021a).

Genetic susceptibility studies utilising the 'heritability' statistic have shown that cognitive abilities, motor attributes, morphological dimensions, functional capacities, and personality traits are moderately to highly hereditary (Georgiades et al., 2017). The heritability estimates for injuries are unclear as the exact aetiology of these multifactorial conditions remains to be elucidated. However, several heritability studies have reported evidence of a genetic component to injuries. For example, Harvie et al. (2004) reported there was a greater risk to siblings suffering a full thickness tear of the rotator cuff, than the spouses of participants who had previously sustained the same injury. Similarly, Flynn et al. (2005) showed that individuals who reported a family history of sustaining an anterior cruciate ligament (ACL) rupture, were twice as likely to suffer an ACL rupture, compared to individuals with no prior

family history of ACL rupture. Moreover, Kraemer et al. (2012) revealed that Achilles tendinopathy occurs significantly more in individuals with a prior family history of sustaining the injury. In addition, a recent study on lifetime ACL rupture risk (involving 88,414 twins) reported a heritability estimate of 69% (Magnusson et al., 2021).

Although heritability studies are important, as they reveal the possibility of a genetic predisposition, they fail to provide information concerning which specific genetic variants (i.e., polymorphisms) are responsible (Guilherme et al., 2014). Therefore, the focus of current football genomic research is on further understanding genotype-phenotype relationships, through the exploration of genetic association. Many empirical studies have investigated genetic association with injury. However, to the author's knowledge, no study has reviewed genetic associations with injury specifically in football players, with the only other team-sport review (non-systematic) having been completed in rugby (Brazier et al., 2019). Therefore, the aim of this review was to synthesise genetic association studies that have investigated injury involving football players to identify the genetic variants which have the most empirical evidence to date.

4.2 Methodology

4.2.1 Search strategy and eligibility criteria

The search strategy followed the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) 2020 guidelines (Page et al., 2021). A comprehensive search of the Pubmed, SPORTDiscus, and MEDLINE databases was conducted up until March 11th 2022, using the Boolean search of: ((football OR soccer) AND (injury OR injuries)) AND (genetics OR gene OR genotype OR snp OR polymorphism)). Additionally, Google Scholar was searched using word combinations of the aforementioned Boolean search. Furthermore, reference lists of the identified articles and reviews were searched for additional relevant studies and forward citation tracking was performed on all eligible studies. Studies were

included if they met the following inclusion criteria: (a) were primary cohort or case-control investigations, (b) reported that football players were included in their population sample, (c) examined the association of a genetic variant with injury, and (d) were published in the English language.

4.2.2 Data extraction and analysis

The following data was extracted from included studies: first author's name and year of publication; number of participants, footballers, and controls; sex; age; nationality; ethnicity; study design; injury phenotype; injury measurement; names of genes and polymorphisms investigated; and, main findings. It was deemed that a statistically pooled quantitative synthesis of the extracted data could not be performed due to the variation in: genes; polymorphisms; injuries; and, ethnicities. Therefore, due to the observed heterogeneity between studies, a narrative synthesis was chosen as the appropriate method with which to summarise results. The methodological quality of the studies was assessed using the Newcastle-Ottawa Scale (NOS; Stang, 2010), as utilised in previous research (Lv et al., 2018; Wang et al., 2017). Studies were scored (0-9 stars) based on the items within the three factors of the appropriate case-control (selection; comparability; exposure) or cohort (selection; comparability; outcome) version of the NOS.

4.3 Results

4.3.1 Search process

The literature search, selection of eligible studies, data extraction, and methodological quality was performed in duplicate and independently by the author and a co-author, culminating in 34 studies being judged as adequately meeting the predetermined inclusion criteria and subsequently being included in the final analysis (see **Figure 4.1**).

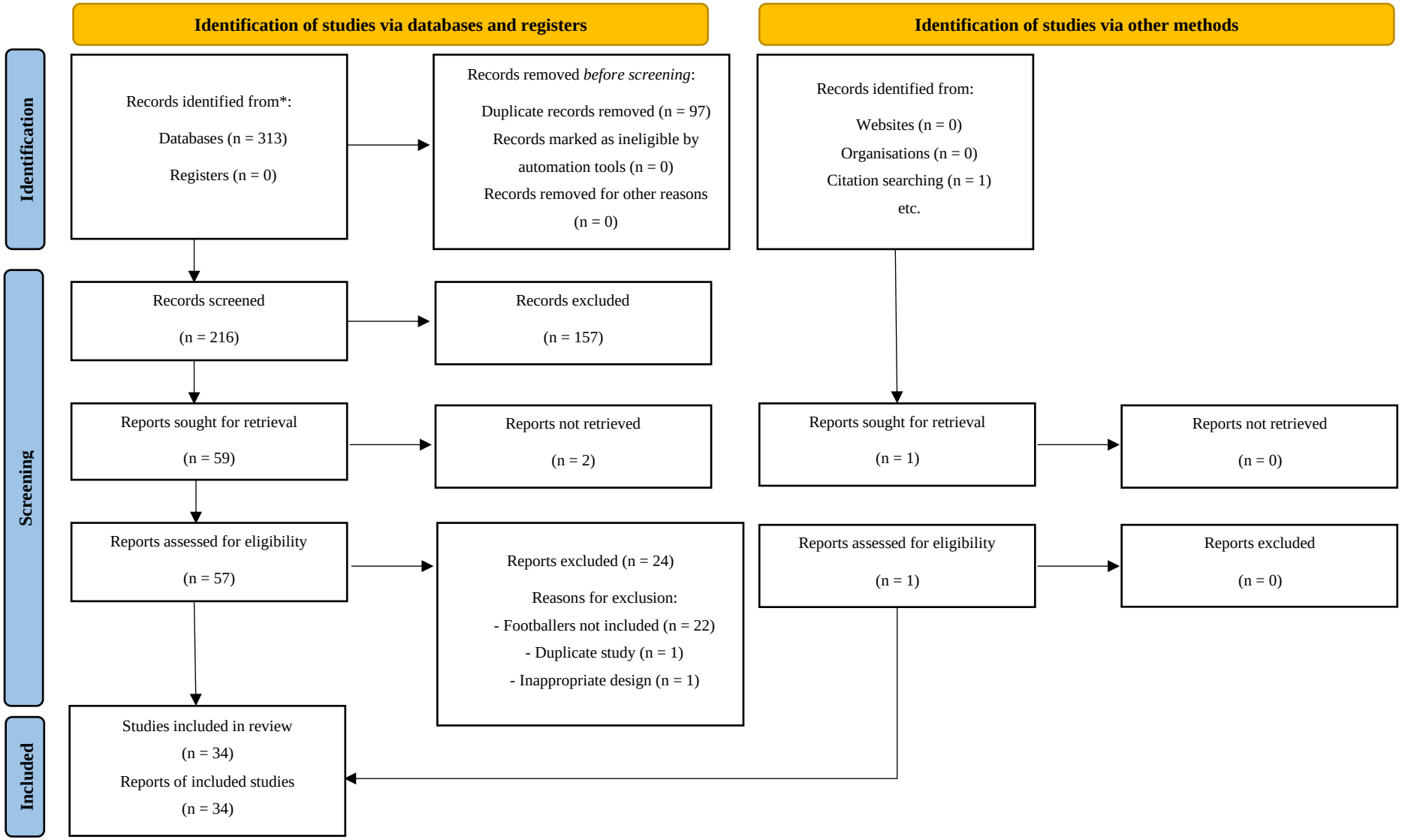


Figure 4.1. Flow diagram of search process.

4.3.2 Study characteristics and quality

Of the 34 included studies, there were 33 candidate gene association studies (CGAS) and one genome-wide association study (GWAS) (see **Table 4.1**). Of which, there were 13 longitudinal studies (ranging from 3 to 10 years follow up), ten cross-sectional studies, ten case-control studies, and one study used both cross-sectional and longitudinal analysis. The total number of participants across all studies was 9642, ranging from 43-1311 participant sample sizes, with a median sample size of 227. Twenty studies included only footballers, whilst the remaining 14 studies included athletes of other sports alongside footballers. Fifteen studies focused exclusively on male subjects, whilst the remaining 19 focused on both male and female subjects. Ethnicity was reported in 25 studies, whilst the remaining nine failed to report ethnicity. Injuries were diagnosed by a medical professional in 27 studies, whilst the injuries in the remaining seven studies were self-reported. The injury phenotypes described in the studies included: ACL rupture (n = 8); concussion (n = 6); musculoskeletal soft tissue injuries (n = 4); muscle injuries (n = 5); stress fracture (n = 3); musculoskeletal injuries (n = 3); hamstring (n = 1); knee and ankle (n = 1); low-back pain (n = 1); medial collateral ligament (MCL; n = 1); tendinopathy (n = 1). The NOS scores of the case-control studies ranged from 4-9, with a mean score of 7.1. The NOS scores of the cohort studies ranged from 3-8, with a mean score of 6.7. The overall mean NOS score across all studies was 6.8 (see **Table 4.2**).

Table 4.1. Study characteristics and main findings.

Study	Participants	Injury	Polymorphism	Main findings
Kristman et al. (2008)	318 college male and female athletes (aged 20.5 ± 2.4 years), including 53 football players.	Concussion	*APOE* E4 isoform (rs429358 + rs7412)	No association between *APOE* E4 and concussion.
Terrel et al. (2008)	195 male college American football players vs 18 male and 18 female college football players (aged 18-30 years).	Self-reported Concussion	*APOE* (rs429358 + rs7412) isoforms (E2, E3, E4); *APOE* (rs405509); *MAPT* (rs2258689, rs10445337)	College athletes may have an increased risk of concussion if they have *APOE* rs405509 TT genotype (OR = 2.7, 95% CI: 1.1-6.8; ***p* = 0.03**).
Tierney et al. (2010)	163 college male American football and 33 college female football athletes (aged 19.7 ± 1.5 years).	Self-reported Concussion	*APOE* (rs429358 + rs7412) isoforms (E2, E4); *APOE* (rs405509)	There was a significant association (***p* = 0.05**; OR = 9.8, CI: 1.00-96.55) between carrying all *APOE* rare alleles (E2, T allele; E4, C allele; rs405509, T allele) and the history of a previous concussion. There was also a significant association (***p* = 0.04**; OR = 8.4, 95% CI: 1.03-68.79) between carrying the *APOE* rs405509 T allele and experiencing 2 or more concussions.
McDevitt et al. (2011)	96 college American football and women's football athletes (aged 19.5 ± 1.4 years).	Self-reported Concussion	*NEFH* (rs165602)	No association between *NEFH* rs165602 and concussion.
Ficek et al. (2013)	91 elite male Polish Caucasian football players (aged 23 ± 3 years) with surgically diagnosed ACL ruptures vs 143 healthy male elite football players (aged 25.2 ± 2.6 years).	ACL rupture (non-contact)	*COL1A1* (rs1107946, rs1800012)	No significant differences in either polymorphism's genotype distributions and allele frequencies between groups. The G-T (rs1107946, rs1800012) haplotype was associated with decreased risk of injury (***p* = .048**).
Pruna et al. (2013)	73 elite male Caucasian (n = 43), Black African (n = 11) and Hispanic (n = 19) football players (aged 19–35 years).	Non-contact musculoskeletal soft tissue injuries	*ELN* (rs2289360); *TTN* (rs2742327); *SOX15* (rs4227); *IGF2* (rs3213221); *CCL2* (rs2857656); *COL1A1* (rs1800012); *COL5A1* (rs12722); TNC (rs2104772)	*IGF2* GC genotype (***p* = 0.034**) was associated with less severe muscle injuries. *CCL2* GG genotype was associated with more severe muscle injuries (***p* = 0.026**). *ELN* AA genotype was associated with more severe ligament injuries (***p* = 0.009**); while the AG genotype was associated with faster recovery time (***p* = 0.043**).
Ficek et al. (2014)	91 elite male Polish Caucasian football players (aged 23 ± 3 years) with surgically diagnosed ACL ruptures vs 143 healthy male elite football players (aged 25.2 ± 2.6 years).	ACL rupture (non-contact)	*COL12A1* (rs970547)	No significant differences in genotype distributions and allele frequencies between groups.

Mannion et al. (2014)	227 Caucasian males and females with surgically diagnosed ACL ruptures (aged 26.8 ± 11.0 years) and 234 males and females without any history of ACL injuries (aged 29.3 ± 11.3 years), including 14 football players.	ACL rupture (non-contact & contact)	*ACAN* (rs2351491, rs1042631, rs1516797); *BGN* (rs1126499, rs1042103); *DCN* (rs13312816, rs516115); *FMOD* (rs7543148, rs10800912); *LUM* (rs2268578)	The G allele of *ACAN* rs1516797 was significantly under-represented in the controls (***p* = 0.024**; OR = 0.72, 95% CI: 0.55-0.96). For *DCN* rs516115, the GG genotype was significantly over-represented in female controls (***p* = 0.015**; OR = 9.231, 95% CI: 1.16-73.01) and the AA genotype was significantly under-represented in controls (***p* = 0.013**; OR = 0.33; 95% CI 0.14-0.78) compared with the female non-contact ACL injury subgroup. *ACAN* haplotype containing alleles T–C–T was significantly over-represented (***p* = 0.001**) in the CON group, while T–C–G was significantly under-represented (***p* = 0.005**) in the CON group. *BGN* C–G haplotype was significantly over- represented (***p* = 0.027**) in the female CON group. *LUM* and *DCN* T–A–G inferred haplotype was significantly over-represented (***p* = 0.038**) in the CON group.
Massidda et al. (2014)	64 Caucasian male professional football players.	Muscle injuries	*COL5A1* (rs12722), *SLC16A1* (rs1049434), *VDR* (rs1544410, rs7975232, rs2228570)	No significant associations were found between injuries and polymorphisms.
Rahim et al. (2014)	227 Caucasian males and females with surgically diagnosed ACL ruptures (aged 26.8 ± 11.0 years) and 234 males and females without any history of ACL injuries (aged 29.3 ± 11.3 years), including 14 football players.	ACL rupture (non-contact & contact)	*VEGFA* (rs699947, rs1570360, rs2010963); *KDR* (rs2071559, rs1870377); *NGFB* (rs6678788); *HIF1A* (rs11549465)	The *VEGFA* rs699947 CC genotype (***p* = 0.010**; OR = 1.92; 95% CI: 1.17–3.17) was significantly over-represented within participants with non-contact ACL ruptures. The *VEGFA* rs1570360 GA genotype was significantly over-represented in the CON group (***p* = 0.007**; OR = 1.70, 95% CI: 1.16–2.50). The *KDR* rs2071559 GA genotype was significantly over-represented in the female controls (***p* = 0.023**; OR = 2.16, 95% CI: 1.11–4.22). *VEGFA* A-A-G haplotype was significantly over- represented (***p* = 0.017**) in the CON group compared to the NON subgroup. *KDR* G-A haplotype was noted to be significantly under-represented in controls compared to the ACL group (***p* = 0.014**), as well as the NON subgroup (***p* = 0.001**). Similarly, the G-A haplotype was significantly under-represented (***p* = 0.035**) in the male controls compared to the NON subgroup.
Massidda et al. (2015a)	54 male elite Caucasian football players (aged 25.9 ± 4.3 years).	Indirect musculoskeletal injuries	*VDR* (rs2228570, rs1544410, rs7975232)	No significant differences were identified in injury incidence and severity between genotypes However, rs7975232 genotypes accounted for 18% of injury severity (***p* = 0.002**). Therefore, rs7975232 genotypes may have an influence on the severity of muscle injury in top-level football players.
Massidda et al. (2015b)	173 Italian Caucasian male elite football players (aged 19.4 ± 5.2 years).	Indirect musculoskeletal injuries	*SLC16A1* (rs1049434)	Significant differences were found between genotypes and incidence of muscle injuries (***p* = 0.048**). Specifically, football players with the TT genotype had lower incidence of muscle injuries compared to AA genotype carriers (***p* = 0.044**). No significant difference was found between genotypes, severity of injury and recovery.

Pruna et al. (2015)	73 elite male Caucasian (n = 43), Black African (n = 11) and Hispanic (n = 19) football players (aged 19–35 years).	Non-contact musculoskeletal soft tissue injuries	*ELN* (rs2289360); *TTN* (rs2742327); *SOX15* (rs4227); *IGF2* (rs3213221); *CCL2* (rs2857656); *COL1A1* (rs1800012); *COL5A1* (rs12722); TNC (rs2104772)	The frequency of the SNPs varied among the three ethnic sub-groups (***p* < 0.0001**). Among Caucasians, a significant relationship was observed between ligament injuries and *ELN* (***p* = 0.001**) and between tendinous injuries and *ELN* (***p* = 0.05**) and IGF2 (***p* = 0.05**). Among Hispanics, there was a significant relation between muscle injuries and *ELN* (***p* = 0.032**) and *IGF2* (***p* = 0.016**).
Varley et al. (2015)	518 elite male and female athletes, including 218 football players. Elite athletes were mainly white Caucasian (83.2% in the stress fracture Cases and 79.9% in the non-stress fracture Controls)	Stress fracture	*RANK* (rs3018362); *RANKL* (rs1021188, rs9594738); OPG (rs4355801)	*RANK* rs3018362 GA + AA genotypes (***p* = 0.049**) and *RANKL* rs1021188 AA genotype (***p* = 0.024**) were associated with stress fracture prevalence. *RANKL* rs1021188 AA genotype (***p* = 0.006**) and *OPG* rs4355801 GA + GG genotypes (***p* = 0.042**) were associated with multiple stress fracture injuries.
Artells et al. (2016)	60 elite European football players (aged 25.52 ± 2.5 years).	MCL	*ELN* (rs2289360)	No significant associations were observed.
Varley et al. (2016)	210 active duty Israel Defence Force soldiers and 518 elite male and female athletes, including 218 football players. Elite athletes were mainly white Caucasian (83.2% in the stress fracture Cases and 79.9% in the non-stress fracture Controls)	Stress fracture	*P2RX7* (rs208294, rs1718119, rs2230912, rs3751143, rs1653624)	The C allele of rs3751143 was associated with stress fracture injury, whilst the A allele of rs1718119 was associated with a reduced occurrence of stress fracture injury in military conscripts (***p* < 0.05**). The C allele of rs3751143 was associated with stress fractures in elite athletes (***p* = 0.05**), whereas the A allele of rs1718119 was associated with reduced multiple stress fracture cases in elite athletes (***p* < 0.01**).
Cauci et al. (2017)	60 Italian Caucasian male and female athletes (aged 33.9 ± 13.3 years), including 7 football players.	Self-reported LBP	*VDR* (rs2228570)	The FF genotype was significantly associated with LBP (***p* = 0.015**; adjusted OR = 5.78, 95% CI: 1.41–23.8). The F allele was a 2-fold risk factor to develop LBP (***p* = 0.046**; adjusted OR = 2.55, 95% CI: 1.02–6.43) while the f allele was protective.
Cięszczyk et al. (2017)	229 elite Polish Caucasian male and female football players who suffered an ACL vs 143 uninjured male and female athletes.	ACL rupture (non-contact)	*ACAN* (rs1516797); *BGN* (rs1042103, rs1126499); *DCN* (rs516115); *VEGFA* (rs699947)	*ACAN* rs1516797 GT genotype was underrepresented in the CON group (***p* = 0.017**; OR = 1.68, 95% CI: 1.09-2.57) when all participants were investigated. *BGN* rs1042103 A allele was significantly under-represented in the male CON group compared to the male ACL group (***p* = 0.029**; OR = 1.5, 95% CI: 1.05-2.15).
Pruna et al. (2017)	74 elite male Spanish football players (aged 19–35 years).	Non-contact musculoskeletal soft tissue injuries	*LIF* (rs929271, rs737812); *CCL2* (rs1860189); *GEFT* (rs11613457); *MYF5* (rs1163263); *DES* (rs60794845, rs58999456); *HGF* (rs5745678, rs5745697, rs1011694); *MMP3* (rs679620); *GDF5* (rs143383)	*HGF* (rs5745678) T allele protects against severe injuries (***p* = 0.002**) and is associated with shorter recovery time (***p* = 0.009**). Injuries associated with the CA/AA genotypes in *HGF* (rs5745697) were not severe (***p* = 0.008**) and resulted in a shorter recovery time than the TT genotype (***p* = 0.020**). Injuries associated with the AT/TT genotypes in *HGF* (rs1011694) were mild or moderate (***p* = 0.019**). Injuries associated with the *GEFT* (rs11613457) GG genotype required less recovery time than the GA genotype (***p* = 0.004**).

Cochrane et al. (2018)	250 collegiate student male and female athletes (aged 19.0 ± 1.3 years), including 35 men's football, and 32 women's football players.	Self-reported Concussion	*APOE* (rs429358 + rs7412) isoforms (E2, E3, E4); *APOE* (rs405509); *COMT* (rs4680); *DRD2* (rs18000497)	No polymorphism significantly predicted concussion history.
Larruskain et al. (2018)	107 elite male Spanish Caucasian football players (aged 20 ± 4 years).	Hamstring (non-contact)	*COL5A1* (rs16399, rs12722)*; COL1A1* (rs1107946, rs1800012)*; CASP8* (rs3834129, rs1045485)*; MMP3* (rs679620)*; MMP1* (rs1799750)*; NOS3* (rs1799983)*; DCN* (rs516115)*; HIF1A* (rs11549465)*; MMP12* (rs2276109)*; ADAM12* (rs3740199)*; SOX15* (rs4227)*; TNC* (rs2104772)*; CCL2* (rs2857656)*; VEGFA* (rs2010963)*; ADAMTS5* (rs226794)*; ACTN3* (rs1815739)*; ACAN* (rs1516797)*; ADAMTS2* (rs1054480)*; IL6* (rs1800795)*; GDF5* (rs143383)*; ACE* (rs1799752)*; COL12A1* (rs970547)*; SOD2* (rs4880)*; MLCK* (rs2700352)*; TIMP2* (rs4789932)*; IL6R* (rs2228145)*; ADAMTS14* (rs4747096)*; EMILIN1* (rs2289360)*; TTN* (rs2742327)*; IGF2* (rs3213221)*; TNF* (rs1800629)*; IL1A* (rs1800587)*; CCR2* (rs768539)*; IL1B* (rs1143634)	Five SNPs were found to be significantly associated with hamstring injury in a multivariable model: *MMP3* rs679620 A vs G (***p* = 0.0000062**; HR = 2.06, 95% CI: 1.51–2.81) *TNC* rs2104772 A vs T (***p* = 0.004**; HR = 1.65, 95% CI: 1.17–2.32) *IL6* rs1800795 GG vs GC + CC (***p* = 0.01**; HR = 1.68, 95% CI: 1.11–2.53) *NOS3* rs1799983 G vs T (***p* = 0.04**; HR = 1.35, 95% CI: 1.01–1.79) *HIF1A* rs11549465 CC vs CT (***p* = 0.05**; HR = 2.08, 95% CI: 1.00–4.29)
Lulińska-Kuklik et al. (2018)	134 elite male Polish Caucasian football players (aged 23.4 ± 3.1 years), with ACL ruptures vs 211 male elite football players (aged 25.3 ± 3.4 years), without any self-reported history of ligament or tendon injury.	ACL rupture (non-contact)	*COL5A1* (rs12722, rs13946)	*COL5A1* rs13946 (CC + CT vs TT) was associated with less injuries (***p* = 0.039**) when a dominant mode of inheritance was tested. C-C haplotype was found to be overrepresented in the control group compared to the ACL group (***p* = 0.038**) when the dominant model was tested.
McCabe & Collins (2018)	289 male football players (aged 18–32 years).	Knee and ankle	*GDF5* (rs143383); *AMPD1* (rs17602729); *COL5A1* (rs12722); *IGF2* (rs680)	*GDF5* TT genotype had more ankle injuries and total injuries (***p* < 0.001**) compared to players with CC and CT genotypes and had more knee injuries (***p* < 0.001**) compared to those in the CT group. *AMPD1* CC genotype had fewer ankle, knee and total injuries (***p* < 0.001**). *COL5A1* TT genotype had more ankle, knee and total injuries (***p* < 0.001**). *IGF2* GG genotype had more knee and total injuries (***p* < 0.001**) compared to players with AA and AG genotypes and had more ankle injuries (***p* < 0.001**) compared to those with AG genotype.

Terrell et al. (2018)	1056 collegiate athletes, including 155 male and female football players, (aged 19.7 ± 1.5 years). Ethnicity - 59.4% Caucasian, 35.0% African-American, and 5.6% other race.	Concussion	*APOE* (rs429358 + rs7412) isoforms (E2, E3, E4); *APOE* (rs405509); *MAPT* (rs2258689, rs10445337); *IL6* (rs1800795); *IL6R* (rs2228145)	Positive association between *IL6R* CC (***p* = 0.001**) and a negative association between *APOE* E4 (***p* = 0.03**) and the risk of concussion. Unadjusted and adjusted logistic regression analysis showed a significant association between *IL6R* CC and concussion (***p* = 0.002**; OR = 3.48, 95% CI: 1.58-7.65) and between *APOE* E4 and concussion (***p* = 0.04**; OR = 0.61, 95% CI: 0.38-0.96).
Varley et al. (2018)	518 elite male and female athletes, including 218 football players (aged 24.2 ± 5.5 years). Athletes were mainly white Caucasian (83.2% in the stress fracture Cases and 79.9% in the non-stress fracture Controls).	Stress fracture	*VDR* (rs1544410, rs731236, rs7975232, rs10735810); *GC* (rs7041, rs4588); *WNT16* (rs3801387); *SOST* (rs1877632)*; COL1A1* (rs1800012)*; LRP5* (rs3736228)*; CTR* (rs1801197)	*SOST* rs1877632 and *VDR* rs10735810 and rs731236 were associated with stress fracture (***p* < 0.05**). In the whole cohort, rs1877632 heterozygotes and homozygotes of the rare allele (T) combined made up 59% of stress fracture sufferers in comparison to 46% in the non-stress fracture group (***p* = 0.05**). In the multiple stress fracture cohort, homozygotes of the rare allele of rs10735810 (f) and rs731236 (T) showed an association with stress fracture when compared to those homozygotes for the common allele combined with heterozygotes (***p* = 0.03**; ***p* = 0.01**).
Clos et al. (2019)	43 elite male Caucasian (53.5%), Black African (16.3%) and Hispanic (30.2%) football players (aged 20-37 years).	Non-contact musculoskeletal soft tissue injuries	*ACTN3* (rs1815739)	*ACTN3* rs1815739 XX genotype was associated with non-contact musculoskeletal soft-tissue injury incidence (***p* = 0.003**). Average number of injuries per season; XX (2.78), RR (1.51) and RX (0.83).
Kumagai et al. (2019)	1311 elite Japanese male and female athletes, including 480 football players (aged 20.6 ± 2.9 years).	Self-reported non-contact muscle injuries	*ESR1* (rs2234693, rs9340799)	Genotype frequencies for *ESR1* rs2234693 CT were significantly different between the injured and uninjured groups in a C-allele dominant CC + CT vs TT (***p* = 0.016**; OR = 0.62, 95% CI: 0.43–0.91; P) and additive CC vs CT vs TT (***p* = 0.008**; OR = 0.70, 95% CI: 0.53–0.91) model in all athletes. Results suggest that the *ESR1* rs2234693 C allele, in contrast to the T allele, provides protection against muscle injury.
Lulińska-Kuklik, Laguette, et al. (2019)	229 elite Polish Caucasian male and female football players who suffered an ACL injury (aged 26 ± 4 years) vs 192 male and female controls (aged 25 ± 3 years).	ACL rupture (non-contact)	*TNC* (rs1330363, rs2104772, rs13321)	Genotype and allele frequencies of *TNC* variants did not differ between cases and controls.
Lulińska-Kuklik, Leźnicka, et al. (2019)	222 elite Polish Caucasian male and female football players who suffered an ACL injury (aged 26 ± 4 years) vs 190 male and female controls (aged 25 ± 3 years).	ACL rupture (non-contact)	*VEGFA* (rs699947, rs1570360, rs2010963)	*VEGFA* rs2010963 was associated with risk of ACL rupture in the codominant (***p* = 0.047**) and recessive model (***p* = 0.017**). In the latter, the CC genotype was overrepresented among individuals with ACL rupture (23.4% vs 14.2%, OR=1.85 [1.11-3.08]).
Massidda et al. (2019)	257 elite male Italian Caucasian football players (aged 21.± 5.3 years) vs 263 untrained males (aged 22.4 ± 6.2 years).	Non-contact muscle injuries	*ACTN3* (rs1815739)	*ACTN3* XX players were more susceptible to injury (***p* = 0.02**; OR = 2.66, 95% CI: 1.09-6.63) and severe injury (***p* = 0.0054**; OR = 2.13, 95% CI: 1.25-3.74) than *ACTN3* RR players.

Rodas et al. (2019)	363 elite male and female team sport athletes, including 223 football players (aged 25 ± 6.5 years). Caucasian (n = 327), African American (n = 21), Brazilian (n = 12), and Asian (n = 3).	Tendinopathy	GWAS - 495,837 initially and 1,419,369 after synthetic variant imputation	Suggestive association ($P < 10^{-5}$) was found for *GJA1* (rs11154027), *VAT1L* (rs4362400), and *CNTP2* (rs10263021). Carriage of the A allele for rs11154027 ($p = 1.01 \times 10^{-6}$) and G allele for rs4362400 ($p = 9.6 \times 10^{-6}$) was associated with a higher risk of tendinopathy. Carriage of the A allele for rs10263021 ($p = 4.5 \times 10^{-6}$) was associated with a lower risk of tendinopathy.
Massidda et al. (2020)	710 male elite football players from Italy (n = 341, aged 19.9 ± 5 years) and Japan (n = 369, aged 20.8 ± 1.4 years).	Non-contact muscle injuries	*ACE* (rs4341)	In the Japanese cohort, the *ACE* I/D polymorphism was significantly associated with muscle injury using the D-dominant model (OR: 0.48, 95% CI: 0.24–0.97, ***p* = 0.04**). The meta-analysis showed that in the pooled model (Italian and Japanese populations), the frequencies of the DD+ID genotypes were significantly lower in the injured groups than in non-injured groups (OR: 0.61, 95% CI: 0.38–0.98, ***p* = 0.04**).
Rodas et al. (2021)	46 male and female players (aged 26.1 ± 4.6 years) from a top-level professional football team.	Non-contact muscle injuries	*ACTN3* (rs1815739)	There was a trend towards a higher risk of muscle injury associated with the XX genotype (P = 0.092, with no injury-free XX player during the 5-year study period) and a significant genotype effect for the time needed to return to play (***p* = 0.044**, with the highest value shown for the XX genotype, i.e., 36 ± 26 days, vs. 20 ± 10 and 17 ± 12 days for RR and RX, respectively).
Hall et al. (2022)	402 Caucasian male academy football players (aged 9-23 years) from England, Spain, Uruguay, and Brazil, whose maturity status was defined as pre- or post-peak height velocity (PHV).	Musculoskeletal injuries	*ACTN3* (rs1815739), *CCL2* (rs2857656), *COL1A1* (rs1800012), *COL5A1* (rs12722), *EMILIN1* (rs2289360), *IL6* (rs1800795), *MMP3* (rs679620), *MYLK* (rs28497577), *VEGFA* (rs2010963)	Pre-PHV *COL5A1* rs12722 CC homozygotes had relatively higher prevalence of any musculoskeletal soft tissue (22.4% vs. 3.0%, ***p* = 0.018**) and ligament (18.8% vs. 11.8%, ***p* = 0.029**) injury than T-allele carriers, while *VEGFA* rs2010963 CC homozygotes had greater risk of ligament/tendon injury than G-allele carriers (***p* = 0.011**). Post-PHV *IL6* rs1800795 CC homozygotes had a relatively higher prevalence of any (67.6% vs. 40.6%, ***p* = 0.003**) and muscle (38.2% vs. 19.2%, ***p* = 0.013**) injuries than G-allele carriers. Relatively more post-PHV *EMILIN1* rs2289360 CC homozygotes suffered any injury than CT and TT genotypes (56.4% vs. 40.3% and 32.8%, ***p* = 0.007**).

Note. ACL = Anterior Cruciate Ligament; MCL = Medial Collateral Ligament; LBP = Low Back Pain; GWAS = Genome-Wide Association Study; OR = Odds Ratio; CI = Confidence Interval.

Table 4.2. Quality assessment of studies.

	Newcastle-Ottowa Scale Score			
Study	Selection	Comparability	Outcome / Exposure	Total
Kristman et al. (2008)	★★★	★★	★★★	8
Terrell et al. (2008)	★★★	★★	★★	7
Tierney et al. (2010)	★★★		★★	5
McDevitt et al. (2011)	★★★★	★★	★★★	9
Ficek et al. (2013)	★★★	★★	★★★	8
Pruna et al. (2013)	★★★★	★	★★★	8
Ficek et al. (2014)	★★★	★★	★★★	8
Mannion et al. (2014)	★★★★		★★	6
Massidda et al. (2014)	★★★★	★	★★★	8
Rahim et al. (2014)	★★		★★	4
Massidda et al. (2015a)	★★★★	★	★★★	8
Massidda et al. (2015b)	★★★★	★	★★★	8
Pruna et al. (2015)	★★★★		★★★	7
Varley et al. (2015)	★★★	★	★★	6
Artells et al. (2016)	★★		★★★	5
Varley et al. (2016)	★★★	★	★★	6
Cauci et al. (2017)	★★	★	★★	5
Cieszczyk et al. (2017)	★★★★	★★	★★	8
Pruna et al. (2017)	★★★★	★	★★★	8
Cochrane et al. (2018)	★		★★	3
Larruskain et al. (2018)	★★★★	★	★★★	8
Lulińska-Kulik et al. (2018)	★★★★	★★	★★	8
McCabe & Collins (2018)	★★		★★★	5
Terrell et al. (2018)	★★★★	★	★★★	8
Varley et al. (2018)	★★★	★	★★	6
Clos et al. (2019)	★★★★	★	★★	7
Kumagai et al. (2019)	★	★	★★	4
Lulińska-Kulik, Laguette, et al. (2019)	★★★★	★★	★★	8
Lulińska-Kulik, Leźnicka, et al. (2019)	★★★★	★★	★★	8
Massidda et al. (2019)	★★★	★	★★★	7
Rodas et al. (2019)	★★★	★	★★★	7
Massidda et al. (2020)	★★★★	★	★★★	8
Rodas et al. (2021)	★★★★		★★	6
Hall et al. (2022)	★★★★	★	★★	7

Note. Selection = 4 items (1 star can be awarded for each item); Comparability = 1 item (2 stars can be awarded); Outcome / Exposure = 3 items (1 star can be awarded for each item). A maximum score of 9 is possible.

Table 4.3. Polymorphism investigations and associations.

Gene	Abbr	Chr	Polymorphism	*N*	+
Collagen type 5 alpha 1 chain	*COL5A1*	9q34.3	rs12722	7	2*
Collagen type I alpha 1 chain	*COL1A1*	17q21.33	rs1800012	6	0
Actinin alpha 3	*ACTN3*	11q13.2	rs1815739	5	2
Apolipoprotein E	*APOE*	19q13.32	E4 rs429358 & rs7412	5	1
Interleukin 6	*IL6*	7p15.3	rs1800795	4	2*
Vascular endothelial growth factor A	*VEGFA*	6p21.1	rs2010963	4	2
Apolipoprotein E	*APOE*	19q13.32	rs405509	4	1
C-C motif chemokine ligand 2	*CCL2*	17q12	rs2857656	4	1
Tenascin C	*TNC*	9q33.1	rs2104772	4	1
Apolipoprotein E	*APOE*	19q13.32	E2 rs429358 & rs7412	4	0
Aggrecan	*ACAN*	15q26.1	rs1516797	3	2
Elastin	*ELN*	7q11.23	rs2289360	3	1
Insulin like growth factor 2	*IGF2*	11p15.5	rs3213221	3	1
Decorin	*DCN*	12q21.33	rs516115	3	1
Growth differentiation factor 5	*GDF5*	20q11.22	rs143383	3	1
Matrix metallopeptidase 3	*MMP3*	11q22.2	rs679620	3	1
Vitamin D receptor	*VDR*	12q13.11	rs7975232	3	1
Vitamin D receptor	*VDR*	12q13.11	rs2228570	3	1
Vascular endothelial growth factor A	*VEGFA*	6p21.1	rs699947	3	1
Apolipoprotein E	*APOE*	19q13.32	E3 rs429358 & rs7412	3	0
SRY-box transcription factor 15	*SOX15*	17p13.1	rs4227	3	0
Titin	*TTN*	2q31.2	rs2742327	3	0
Vitamin D receptor	*VDR*	12q13.11	rs1544410	3	0
Biglycan	*BGN*	Xq28	rs1042103	2	1
Elastin microfibril interfacer 1	*EMILIN1*	2p23.3	rs2289360	2	1
Hypoxia inducible factor 1 subunit alpha	*HIF1A*	14q23.2	rs11549465	2	1
Interleukin 6 receptor	*IL6R*	1q21.3	rs2228145	2	1
Solute carrier family 16 member 1	*SLC16A1*	1p13.2	rs1049434	2	1
Vascular endothelial growth factor A	*VEGFA*	6p21.1	rs1570360	2	1
Biglycan	*BGN*	Xq28	rs1126499	2	0
Collagen type I alpha 1 chain	*COL1A1*	17q21.33	rs1107946	2	0
Collagen type XII alpha 1 chain	*COL12A1*	6q13-q14.1	rs970547	2	0
Microtubule associated protein tau	*MAPT*	17q21.31	rs10445337	2	0
Microtubule associated protein tau	*MAPT*	17q21.31	rs2258689	2	0
Adenosine monophosphate deaminase 1	*AMPD1*	1p13.2	rs17602729	1	1
Angiotensin I converting enzyme	*ACE*	17q23.3	rs4341	1	1
Collagen type 5 alpha 1 chain	*COL5A1*	9q34.3	rs16399	1	1
Estrogen receptor 1	*ESR1*	6q25.1-q25.2	rs2234693	1	1
Hepatocyte growth factor	*HGF*	7q21.11	rs1011694	1	1
Hepatocyte growth factor	*HGF*	7q21.11	rs5745697	1	1
Hepatocyte growth factor	*HGF*	7q21.11	rs5745678	1	1

Insulin like growth factor 2	*IGF2*	11p15.5	rs680	1	1
Kinase insert domain receptor	*KDR*	4q12	rs2071559	1	1
Nitric oxide synthase 3	*NOS3*	7q36.1	rs1799983	1	1
Purinergic receptor P2X 7	*P2RX7*	12q24.31	rs1718119	1	1
Purinergic receptor P2X 7	*P2RX7*	12q24.31	rs3751143	1	1
Rho guanine nucleotide exchange factor 25	*GEFT*	12q13.3	rs11613457	1	1
Sclerostin	*SOST*	17q21.31	rs1877632	1	1
TNF receptor superfamily member 11a	*RANK*	18q21.33	rs3018362	1	1
TNF receptor superfamily member 11b	*OPG*	8q24.12	rs4355801	1	1
TNF superfamily member 11	*RANKL*	13q14	rs1021188	1	1
Vitamin D receptor	*VDR*	12q13.11	rs731236	1	1
Vitamin D receptor	*VDR*	12q13.11	rs10735810	1	1
ADAM metallopeptidase domain 12	*ADAM12*	10q26.2	rs3740199	1	0
ADAM metallopeptidase with thrombospondin type 1 motif 14	*ADAMTS14*	10q22.1	rs4747096	1	0
ADAM metallopeptidase with thrombospondin type 1 motif 2	*ADAMTS2*	5q35.3	rs1054480	1	0
ADAM metallopeptidase with thrombospondin type 1 motif 5	*ADAMTS5*	21q21.3	rs226794	1	0
Aggrecan	*ACAN*	15q26.1	rs1042631	1	0
Aggrecan	*ACAN*	15q26.1	rs2351491	1	0
Angiotensin I converting enzyme	*ACE*	17q23.3	rs1799752	1	0
C-C motif chemokine ligand 2	*CCL2*	17q12	rs1860189	1	0
C-C motif chemokine receptor 2	*CCR2*	3p21.31	rs768539	1	0
Calcitonin receptor	*CTR*	7q21.3	rs1801197	1	0
Caspase 8	*CASP8*	2q33.1	rs1045485	1	0
Caspase 8	*CASP8*	2q33.1	rs3834129	1	0
Catechol-O-methyltransferase	*COMT*	22q11.21	rs4680	1	0
Decorin	*DCN*	12q21.33	rs13312816	1	0
Desmin	*DES*	2q35	rs58999456	1	0
Desmin	*DES*	2q35	rs60794845	1	0
Dopamine receptor D2	*DRD2*	11q23.2	rs18000497	1	0
Estrogen receptor 1	*ESR1*	6q25.1-q25.2	rs9340799	1	0
Fibromodulin	*FMOD*	1q32.1	rs10800912	1	0
Fibromodulin	*FMOD*	1q32.1	rs7543148	1	0
GC vitamin D binding protein	*GC*	4q13.3	rs7041	1	0
GC vitamin D binding protein	*GC*	4q13.3	rs4588	1	0
Interleukin 1 alpha	*IL1A*	2q14.1	rs1800587	1	0
Interleukin 1 beta	*IL1B*	2q14.1	rs1143634	1	0
Kinase insert domain receptor	*KDR*	4q12	rs1870377	1	0
LDL receptor related protein 5	*LRP5*	11q13.2	rs3736228	1	0
LIF interleukin 6 family cytokine	*LIF*	22q12.2	rs737812	1	0
LIF interleukin 6 family cytokine	*LIF*	22q12.2	rs929271	1	0
Lumican	*LUM*	12q21.33	rs2268578	1	0
Matrix metallopeptidase 1	*MMP1*	11q22.2	rs1799750	1	0
Matrix metallopeptidase 12	*MMP12*	11q22.2	rs2276109	1	0

Myogenic factor 5	*MYF5*	12q21.31	rs1163263	1	0
Myosin light chain kinase	*MLCK/MYLK*	3q21.1	rs2700352	1	0
Myosin light chain kinase	*MLCK/MYLK*	3q21.1	rs28497577	1	0
Nerve growth factor	*NGFB*	1p13.2	rs6678788	1	0
Neurofilament heavy	*NEFH*	22q12.2	rs165602	1	0
Purinergic receptor P2X 7	*P2RX7*	12q24.31	rs208294	1	0
Purinergic receptor P2X 7	*P2RX7*	12q24.31	rs2230912	1	0
Purinergic receptor P2X 7	*P2RX7*	12q24.31	rs1653624	1	0
Superoxide dismutase 2	*SOD2*	6q25.3	rs4880	1	0
Tenascin C	*TNC*	9q33.1	rs13321	1	0
Tenascin C	*TNC*	9q33.1	rs1330363	1	0
TIMP metallopeptidase inhibitor 2	*TIMP2*	17q25.3	rs4789932	1	0
TNF superfamily member 11	*RANKL*	13q14	rs9594738	1	0
Tumor necrosis factor	*TNF*	6p21.33	rs1800629	1	0
Wnt family member 16	*WNT16*	7q31.31	rs3801387	1	0

Note. Abbr = Abbreviation; Chr = Chromosome location; *N* = Number of studies; + = Association; * = Contrasting allelic associations.

4.3.3 Genetic variants

Across the 33 CGASs, a total of 99 unique polymorphisms were assessed within 63 genes (see **Table 4.3**). Forty-one unique polymorphisms were associated with injury at least once, whereas three polymorphisms had their specific allelic associations with injury replicated at least twice in independent cohorts: Actinin alpha 3 (*ACTN3*; rs1815739), Aggrecan (*ACAN*; rs1516797), and Vascular endothelial growth factor A (*VEGFA*; rs2010963). More specifically, the X/X genotype of *ACTN3* (rs1815739) was associated with an increased susceptibility to non-contact muscle injuries (Clos et al., 2019; Massidda et al., 2019). Whereas the G allele and T/T genotype of *ACAN* (rs1516797) were associated with increased and decreased susceptibility to ACL injuries, respectively (Mannion et al., 2014; Cięszczyk et al., 2017), whilst the C/C genotype of *VEGFA* (rs2010963) was associated with an increased risk of ACL rupture (Lulińska-Kuklik, Leźnicka, et al., 2019) and ligament or tendon injuries (Hall et al., 2022). In the only GWAS (Rodas et al., 2019), three additional polymorphisms had a 'suggestive' association ($p < 10^{-5}$) with tendinopathy; however, they failed to reach the study's set ($p < 10^{-7}$) and typical ($p < 10^{-8}$) genome-wide statistical significance thresholds.

4.4 Discussion

The aim of this review was to synthesise genetic association studies that have investigated injury involving football players to identify which genetic variants have the most empirical evidence to date. To the author's knowledge, this is the first review to explore this within a football context. The main findings of this review show that of the 99 unique polymorphisms that have been assessed regarding genetic associations with injury in football players, only the *ACTN3* (rs1815739), *ACAN* (rs1516797), and *VEGFA* (rs2010963) polymorphisms presented similar findings in independent cohorts. Replication is vitally important in genetic association research, as associations in preliminary studies are often overstated. Indeed, a meta-analysis showcased that it is common for subsequent studies to report more modest associations, compared to superior associations of initial studies when investigating novel genetic variants (Ioannidis et al., 2001). As such, preliminary reported genetic associations should be carefully interpreted until a subsequent study using an independent population sample replicates the results. Thus, as *ACTN3* (rs1815739), *ACAN* (rs1516797), and *VEGFA* (rs2010963) were the only three isolated polymorphisms that appeared to replicate their specific allelic associations in more than one independent cohort, they will be discussed in greater detail. Furthermore, a critical evaluation of the intra- and inter-study methodological limitations will be provided to examine the reliability of their individual findings and validity of their replications.

The R allele of *ACTN3* (rs1815739) is regarded as beneficial to strength/power performance and in a recent meta-analysis has been associated with professional status within football (McAuley, Hughes, et al., 2021b). Regarding injury susceptibility, Moreno et al. (2020) recently reported that endurance runners with *ACTN3* X/X genotype were at an increased risk of sustaining a muscle injury. One of the functions of *ACTN3* involves encoding the actinin alpha 3 protein (North et al., 1999), which is a vital structural component of the Z-line as it, along with actin-containing filaments, anchors and stabilises the muscle contractile

mechanism (Mills et al., 2001). The rs1815739 polymorphism can produce a deficient protein when a premature stop codon (X) replaces arginine (R) at residue 577 (Yang et al., 2003). As such, it is speculated that because X/X individuals have a lack of actinin alpha 3 protein, this reduces the capacity of the skeletal muscle to tolerate the consequent muscle contractions from long-term and exhaustive exercise that facilitate muscle injury (Baltazar-Martins et al., 2020). Specifically, a X/X genotype may create a less powerful link between the actin filaments and the Z-line, which then results in a structural deficiency, leading to a sarcomere more prone to suffering damage under high mechanical stress (Baltazar-Martins et al., 2020). Furthermore, it has been reported that actinin alpha 3 deficient individuals appear to sustain greater muscle damage following physical activity (Lee et al., 2016; Pimenta et al., 2012), but require less recovery time (Coelho et al., 2019; Lee et al., 2016). Therefore, it has also been speculated that although football players possessing a X/X genotype may recover at a faster rate, they are perhaps exposed to an intolerable amount of muscle damage, inhibiting the recovery process (Clos et al., 2019). Indeed, several studies found that X/X homozygotes sustained more severe injuries or were absent for a greater number of days following injury (Hall et al., 2022; Massidda et al., 2019; Rodas et al., 2021). A mechanistic explanation for an increased injury risk in those with the X/X variant has been proposed. The X/X genotype may result in an enhanced activation of calcineurin and a consequent shift in fast-twitch fibres toward oxidative metabolism (Seto et al., 2013). This inter-genotype metabolic handling resulted in significantly higher calcium release during muscle contractions in *ACTN3* knockout mice (Quinlan et al., 2010; Head et al., 2015). Moreover, an additional by-product of complete ACTN3 deficiency is the upregulation and accumulation of other Z-line proteins (Seto et al., 2011). Therefore, in X-allele carriers, this may decrease the stability and rigidity of type IIa fibres (Broos et al., 2012), which may facilitate a greater susceptibility to muscle injuries (Baumert et al., 2019). It is important to note that these suggestions remain speculative, as the exact mechanism to

explain the potential higher incidence of sports-related muscle injury in X/X athletes is yet to be established (Baltazar-Martins et al., 2020).

The *ACAN* (rs1516797) polymorphism was investigated in three independent population samples. Whilst only two of the three investigations found a significant association with injury, it is important to note that the study that showed no significant differences was only investigating associations with muscular (hamstring) injury (Larruskain et al., 2018). Whereas, the two studies that found a significant association with *ACAN* (rs1516797) and injury were investigating ligament (more specifically ACL) injuries. The role of ACAN in the structural governance of the ligament (Young et al., 2011), may provide an explanation for the findings shown. Both studies reported that individuals possessing a G allele of ACAN are more at risk of sustaining an ACL injury, with the T/T genotype being associated with increased protection. The genotype distribution of individuals who sustained an ACL injury was similar between studies (i.e., G/G = 11%, T/G = 47-49%; T/T = 40-42%). The *ACAN* gene encodes the aggrecan protein, which is a large structural proteoglycan mostly abundant in cartilage (Young et al., 2011). Proteoglycans perform a synergistic role in fibrillogenesis, potentially through many of their direct/indirect interactions with several proteins; including, the collagen network and cell-signalling molecules within the extracellular matrix (Heinegård, 2009). The alteration of collagen fibril properties may change various biomechanical and functional components of the ligament, possibly increasing injury risk. Indeed, lower levels of proteoglycan and glycosaminoglycan have been observed in ruptured versus non-ruptured human ACL tissue (Young et al., 2011). As such, proteoglycan encoding genes are deemed viable candidates worthy of investigation regarding associations with ligament injuries. However, the specific biological functions of the T/G genotypes within the *ACAN* (rs1516797) polymorphism are yet to be determined (Heinegård, 2009; Mannion et al., 2014). Until an exact mechanistic explanation for *ACAN* (rs1516797) is determined, it is currently unclear how the shown

associations can be used in the identification of genetic risk factors for injury in football players.

The *VEGFA* (rs2010963) polymorphism was investigated in four studies using independent cohorts. The two studies that reported an association between *VEGFA* (rs2010963) and injury found associations with ACL ruptures (Lulińska-Kuklik, Leźnicka, et al., 2019) and ligament or tendon injuries (Hall et al., 2022). The two studies that reported no association between *VEGFA* (rs2010963) and injury were investigating ACL ruptures (Rahim et al., 2014) and hamstring injuries (Larruskain et al., 2018). The *VEGFA* gene encodes the vascular endothelial growth factor-A protein, which is considered the dominant inducer of angiogenesis (Schneider et al., 2008). Angiogenesis can be described as the formation of new capillary blood vessels from existing micro vessels, which has an important function in numerous biological processes (e.g., embryological development, inflammation, and wound healing) and the pathogenesis of several diseases (e.g., cancer, diabetic retinopathy, and rheumatoid arthritis) (Zhao et al., 2016). Functional polymorphisms of *VEGFA* can alter gene expression and protein production, an imbalance of which can have negative physiological consequences. For instance, it has been reported that an increase in VEGFA expression upregulates the expression of matrix metalloproteinases, which may adversely alter the biomechanical properties of ligaments via compromised extracellular matrix homeostasis (Wang & Keiser, 1998). The C/C genotype of *VEGFA* (rs2010963) has been associated with enhanced protein expression and plasma VEFGA concentration (Schneider et al., 2008). As such, the increased susceptibility of C/C homozygotes to ligament injuries in football may be due to these biological mechanisms associated with increased VEFGA expression.

When critically analysing the methodological approach and cohort characteristics of each *ACTN3*, *ACAN*, and *VEGFA* study, it is clear that there are some significant within- and between-study limitations and variability. As such, this undermines the reliability and validity

of the reported associations. For instance, with regards to the *ACTN3* studies, Massidda et al. (2019) reported on 169 Caucasian male football players of varying ages and competitive playing levels recorded via injury incidence (i.e., total injuries per 1000 hours). Whereas Clos et al. (2019) reported on 43 senior professional male football players of different ethnicities (Caucasian = 23; Hispanic = 13; Black African = 7) recorded via injury rate (i.e., total injuries per season). With regards to the *ACAN* studies, Cięszczyk et al. (2017) reported on a sample of 229 male (n = 158) and female (n = 71) Polish football players of a similar age but varying competitive playing levels who sustained an ACL injury via non-contact mechanisms. Whereas Mannion et al. (2014) reported on a sample of 227 male (n = 166) and female (n = 61) Caucasian athletes from multiple sports (football players = 14 males) of varying age and competitive playing levels who sustained an ACL injury via non-contact (n = 126) and contact mechanisms (n = 101). Finally, with regards to the *VEGFA* studies, Lulińska-Kuklik, Leźnicka, et al's (2019) sample consisted of 222 senior Polish Caucasian male (n = 156) and female (n = 66) football players of a similar age but varying competitive playing levels who sustained an ACL injury via non-contact mechanisms. Whereas Hall et al's (2022) sample consisted of 402 Caucasian male academy football players analysed separately based on maturity status who sustained an injury via contact and non-contact mechanisms. There are some evident issues present in these studies (e.g., small sample sizes, cohort heterogeneity, and population stratification), which are prevalent throughout all studies in this review and are discussed in more detail below (see limitations section). However, a specific limitation is the combination of male and females in regards to ACL injury risk. It is well reported that females are at a higher risk of ACL injury compared to males, with female football players in particular at a 2- to 3-fold higher risk than their male counterparts (see Waldén et al., 2011 for a review). As such, when factoring in these limitations and the considerable between-study variability, the

evidence associating *ACTN3* (rs1815739), *ACAN* (rs1516797), and *VEGFA* (rs2010963) with injury risk in football loses its credibility.

Perhaps the most surprising result of this review was the lack of association and/or replication between some of the most heavily researched genes regarding injury susceptibility in sport; namely, the collagen type 5 alpha 1 chain (*COL5A1*) and collagen type 1 alpha 1 chain (*COL1A1*) genes. Both genes are involved in providing instructions for making components of collagen (Lv et al., 2018; Wang et al., 2017). The genetic variants of both genes have been extensively researched in a variety of sports, particularly the *COL5A1* (rs12722) and *COL1A1* (rs1800012) polymorphisms. Indeed, two recent meta-analyses have reported that individuals possessing the T/T genotype of the *COL5A1* (rs12722) polymorphism are predisposed to a higher risk of sustaining tendon and ligament injuries (Lv et al., 2018); whilst, individuals possessing the T/T genotype of the *COL1A1* (rs1800012) polymorphism have a reduced risk of tendon and ligament injuries (Wang et al., 2017). Within this review, the *COL5A1* (rs12722) polymorphism was studied the most frequently (seven studies), followed by the *COL1A1* (rs1800012) polymorphism (six studies). However, whilst the *COL5A1* (rs12722) polymorphism was found to be associated with injury in two independent cohorts, contrasting allelic associations were reported in each study. More specifically, McCabe & Collins (2018) showed that the T allele was associated with knee and ankle injuries in senior players, whereas Hall et al. (2022) presented associations between the C allele and musculoskeletal soft-tissue injuries and ligament injuries in youth players. As such, with regards to football players, it would appear that the *COL5A1* (rs12722) and *COL1A1* (rs1800012) polymorphisms are not independently associated with injury. However, as also previously mentioned regarding the *ACAN* gene, not all studies investigated the *COL5A1* (rs12722) and *COL1A1* (rs1800012) polymorphisms with the most appropriate soft tissue injury (i.e., tendon and ligaments). Indeed, a number of studies instead chose to focus on muscular injuries, injuries as a whole, and/or

contact injuries. It is important to note the substantial variation in the underpinning mechanism(s) contributing to different injuries and that different injuries are likely to have dissimilar genetic predispositions. Moreover, contact injuries are likely to be less influenced by genetic predispositions and more a consequence of the contributing environmental factors. Consequently, the ability to draw valid conclusions based on genetic associations with contact injuries is extremely problematic. As such, the true association of the *COL5A1* (rs12722) and *COL1A1* (rs1800012) polymorphisms with injury in footballers remains unclear.

4.4.1 Limitations

Several methodological limitations exist across genetic association research investigating injury susceptibility in football. Thus, evaluations of many potential associations between polymorphisms and injury remain unclear and are currently inconclusive. Firstly, many of the studies were conducted using retrospective designs, with participants who self-reported their injuries. This is problematic, as many athletes have insufficient knowledge of what constitutes specific injuries and also allows for the possibility of recall bias (Fuller et al., 2006). Secondly, the sample size of most studies was very small. Consequently, many studies may have lacked the statistical power required to detect a significant association between a polymorphism and an injury, which may partly explain why most of the significant associations reported in this review were in studies with larger sample sizes. It is well known that a study investigating the association of a single polymorphism usually requires a sample size of hundreds, and in some cases, thousands (Hong & Park, 2012). This is because individual polymorphisms most likely only contribute a small amount to injury susceptibility, given that injury is such a multifactorial phenomenon (McCrea et al., 2017). It is also important to note that few studies attempted to investigate the combined influence of multiple polymorphisms with injury risk. Possible gene-gene/gene-environment interactions are crucial regarding injury risk, in order to understand the complex mechanisms and polygenic nature underpinning each condition (McCrea et al., 2017).

Furthermore, the authors discovered only one GWAS within the literature. A GWAS is one of the most appropriate methods with which to assess genetic associations, given their ability to assess over 1,000,000 polymorphisms at once and identify novel genetic associations (Guilherme et al., 2014). As such, it is important that more advanced genomic technology is utilised in the future for studies that are appropriately designed for such.

The heterogeneity of population samples via population stratification is also of concern. Individuals from different ancestral backgrounds inherit distinct allele frequencies. Thus, inter-individual ethnic variation is problematic and can result in spurious associations (Freedman et al., 2004). However, many studies in this review have investigated footballers of diverse ethnicities together. Moreover, several studies have analysed participants of different sexes, playing levels, and sports together. This is problematic, as training and match loads differ across these population substructures, which consequently results in diverse injury susceptibility and further exacerbates sample heterogeneity (Lv et al., 2018). As such, it is essential that future research is conducted with improved methodological approaches and study designs (McAuley, Baker, et al., 2022). Recognisably, it is difficult to obtain a large homogenous cohort of athletes and afford the use of advanced genomic technology (Herbert et al., 2019; McAuley, Baker, et al., 2021; McAuley, Hughes, et al., 2022). Therefore, the participation of organisations in international consortia is imperative (McAuley, Hughes, et al., 2021a; Pitsiladis et al., 2016). This will allow the sharing of data and resources, facilitate superior statistical power, and progress our understanding of the exact underlying pathobiology of injury susceptibility. Finally, it is also important to note that this review in itself is not without its limitations. Specifically, the inclusion of papers published only in the English language at the searching, screening, and analysis phases and the inclusion of solely published papers in general. As such, language restriction and publication bias may have had a significant

effect on the number of studies included in this review, and consequently the number of positive and/or negative associations.

4.4.2 Practical Applications

The utilisation of genetic information to aid in tailoring individualised training program designs for prehabilitation, and the recovery of players susceptible to injury, is an exciting prospect. However, at present, the evidence base supporting the use of genetic testing as a prognostic or diagnosis tool for injury risk in football is weak, and confounded by methodological limitations and inconsistencies. Future research via collaborations with improved methodological approaches and mechanistic studies will help identify significant biological pathways underpinning injury risk. Although, to truly have meaningful clinical application, genetic information will have to be used collectively alongside the various other intrinsic and extrinsic risk factors of injury (e.g., sex, age, injury history, anthropometrical differences, technique, equipment, and player loads) (López-Valenciano et al., 2020; McCall et al., 2015; Vlahovich et al., 2017).

4.5 Conclusion

Currently, 41 polymorphisms have been associated with injury in football at least once, whereas three polymorphisms (i.e., *ACTN3* rs1815739, *ACAN*; rs1516797, and *VEGFA* rs2010963) have had their specific allelic associations with injury replicated at least twice in independent cohorts. However, there are several methodological issues (e.g., small sample sizes, cohort heterogeneity, and population stratification) that limit the subsequent reliability and external validity of findings. As such, within a football context, based on this review, there are currently no replicated and validated genetic variants that warrant the utilisation of genetic information as a prognostic or diagnosis tool for injury risk. The future participation of organisations in international consortia is suggested to combat the current methodological

issues discussed within this review and subsequently improve clarity concerning the underlying genetic contribution to injury susceptibility.

5 Genetic testing in professional football: Perspectives of key stakeholders

McAuley, A. B. T., Hughes, D. C., Tsaprouni, L. G., Varley, I., Suraci, B., Roos, T. R., Herbert, A. J., & Kelly, A. L. (2022e). Genetic testing in professional football: Perspectives of key stakeholders. *Journal of Science in Sport and Exercise*, *4*(1), 49–59. https://doi.org/10.1007/s42978-021-00131-3

Abstract

Genetic research in football is currently in its infancy but is growing rapidly. However, the practical application of genetic testing in football and the views concerning its use are unknown. Thus, the purpose of this study was to assess the current practical application of genetic testing in professional football and provide an insight into the perspectives of key stakeholders (i.e., coaches, practitioners, players). In total, 122 participants completed an online anonymous survey. This consisted of 21 multiple choice and Likert scale questions, with the option of providing an explanation for each response. Findings revealed genetic testing is rarely utilised by key stakeholders (10%) or their respective organisations (14%). However, three quarters (75%) had the opinion that genetic testing will have great utility in the future. The majority (72%) believed genetic testing should be used for athlete development and injury risk, whilst 35% believed that genetic testing should be utilised for talent identification purposes. However, most key stakeholders viewed their own (89%) and their colleagues' (79%) knowledge related to genetic testing as insufficient; mainly due to ineffective current communication methods (91%). Most believed educational workshops are required (71%), whilst nearly all (91%) were interested in developing their expertise on the utility of genetic testing. Genetic testing is rarely used within professional football, although key stakeholders anticipate that it will be utilised

more in the future. As such, educational support may prove valuable in improving key stakeholder knowledge and the practical application of genetic testing in professional football.

5.1 Introduction

Achieving elite status in professional sport is a multifactorial process (Kelly & Williams, 2020). More specifically, task constraints (e.g., deliberate practice; deliberate play), performer constraints (e.g., psychological characteristics; physiological factors), environmental constraints (e.g., relative age effects; birth place effects), and genetic factors have previously been shown to interact to facilitate sporting success at adulthood (Davids & Baker, 2007; Rees et al., 2016). To what extent each of these facets influences performance specifically in football remains unclear (Sarmento et al., 2018). Current research has estimated that the genetic contribution (i.e., heritability) to overall athletic status is ~66% (De Moor et al., 2007), with the estimated genetic influence on specific performance traits ranging broadly from 30-80% (Yan et al., 2016). Moreover, several genetic markers have been identified that may be associated with athlete status and specific performance-related phenotypes in football, such as the alpha-actinin-3 (*ACTN3*) R577X and angiotensin-I-converting enzyme (*ACE*) I/D polymorphisms (see McAuley, Hughes, et al., 2021b for a review).

In light of this potential for genetic variation to influence performance, both direct-to-consumer (DTC) and provider based genetic testing services are now offered by many companies (Webborn et al., 2015). Several DTC companies specifically target professional sport with advertising campaigns, which claim to provide personalised training and nutritional information to optimise performance and reduce injury susceptibility based on an athlete's genotype (Goodlin et al., 2015a). Despite genetic testing in sport being a relatively new field, several athletes, practitioners, and organisations at the professional level have begun to embrace genetic testing as part of their training regime (Goodlin et al., 2015b). Indeed, genetic testing related to sport performance and injury susceptibility has been utilised within several sports in the UK (Varley et al., 2018). Furthermore, genetic testing in sport has been used in a

variety of circumstances throughout the world, such as identifying age, verifying sex, detecting doping, and revealing medical conditions (Patel & Varley, 2019).

Interestingly, there is anecdotal evidence in football that suggests genetic testing is being used to identify and select talented performers (Scott & Kelso, 2008). This is concerning for many researchers and practitioners, not only because the existing evidence which DTC companies base their recommendations on is limited (Tanisawa et al., 2020), but also due to the accompanying social, ethical, and legal issues associated with potential genetic discrimination (Goodlin et al., 2015a). As such, several scientific consensus statements have deemed that the utilisation of genetic information, particularly for predicting future performance, is inappropriate and without scientific creditability (Vlahovich et al., 2017; Webborn et al., 2015). However, as genetic testing in sport is still in its infancy, there is currently limited formal regulation and legal legislation (Patel & Varley, 2019). As a result, organisations within football currently have little guidance on best practices, which may result in key stakeholders (i.e., coaches, practitioners, players) becoming vulnerable to misinformation (Tanisawa et al., 2020).

Despite anecdotal evidence, it is not yet known the extent to which genetic testing is taking place in football and why key stakeholders may or may not use genetic testing. To the authors' knowledge, there are only two peer-reviewed studies that have assessed the current use and opinions of genetic testing in sport (Pickering & Kiely, 2021; Varley et al., 2018). However, these studies only included 23 and 22 stakeholders employed in football respectively, limiting the application of their results to a football-specific context. As such, the aim of this study was to assess the current practical application of genetic testing in professional football by providing an insight into the perspectives of key stakeholders.

5.2 Methodology

5.2.1 Recruitment

Key stakeholders employed in professional football were contacted via email and by word of mouth from pre-existing personal and professional contacts. They were each invited to participate in this current study by completing the online anonymous survey and distribute it to other relevant parties. The study was also posted and advertised on various social media platforms. Invitations and posts included a link to the survey, whereby upon clicking the link, individuals would be subsequently directed to: (a) an information sheet detailing the survey's purpose and eligibility requirements, and (b) an informed-consent form. The inclusion criteria consisted of: (a) being aged at least 18 years, (b) employed as a member of staff at a professional football club or organisation involving player development or contracted as a current player, and (c) providing informed consent. Ethical approval was granted by Birmingham City University via the Health, Education, and Life Sciences Academic Ethics Committee.

5.2.2 Survey

The survey was completed anonymously utilising an online survey tool (https://www.onlinesurveys.ac.uk), which is fully compliant with UK Data Protection laws and meets UK accessibility requirements. The survey comprised of 21 multiple choice and Likert scale questions (e.g., strongly disagree, disagree, unsure, agree, strongly agree; no influence, slightly, unsure, moderately, largely) regarding genetic association research and genetic testing in sport (see **Appendix A** for survey template). Respondents were also offered the opportunity to provide a qualitative explanation for their answers to enrich the quantitative data collected from the questions. Indeed, mixed-methodologies are encouraged in contemporary sport science research to ensure that findings are grounded in participants' real-life experiences (Kelly & Turnnidge, 2021). The questions were broadly separated into the following six

themes: (a) demographics, (b) utilisation, (c) awareness, (d) impact, (e) implementation, and (f) education. Data analysis consisted of frequency-based descriptive analysis, which was provided by the survey software directly, then exported for confirmation and further analysis. Methodological procedures are in accordance with Varley et al. (2018) and Pickering and Kiely (2021).

5.3 Results

5.3.1 Demographics

In total, 122 key stakeholders completed the survey between July 13th, 2020 and September 4th, 2020. The key stakeholders were predominately male (90%) and aged between 25-34 years (48%). Their specific employment roles included: Coach (*n* = 40), Player (*n* = 24), Physiotherapist (*n* = 16), Sport Scientist (*n* = 16), Strength and Conditioning Coach (*n* = 9), Performance Analyst (*n* = 5), and Nutritionist (*n* = 2). The remaining ten stakeholders specified their employment role as "Other", which included: Football Consultant, Football Administrator, Hospitality and Sponsorship Manager, Manager, Operations Manager, Operations Manager of Charitable Projects, Performance Lead, Player Recruitment, Sponsorship Development and Management, and Sport Psychologist. There was a relatively even distribution between those working in Youth/Academy (52%) and Senior (48%) football (see **Figure 5.1**). The key stakeholders represented a wide spectrum of competitive playing standards at each level (e.g., Youth/Academy: International [*n* = 4], Academy Categories 1-4 [*n* = 45], and Non-Academy [*n* = 6]; Senior: International [*n* = 6], Divisions 1-4 [*n* = 32], and Non-League [*n* = 17]).

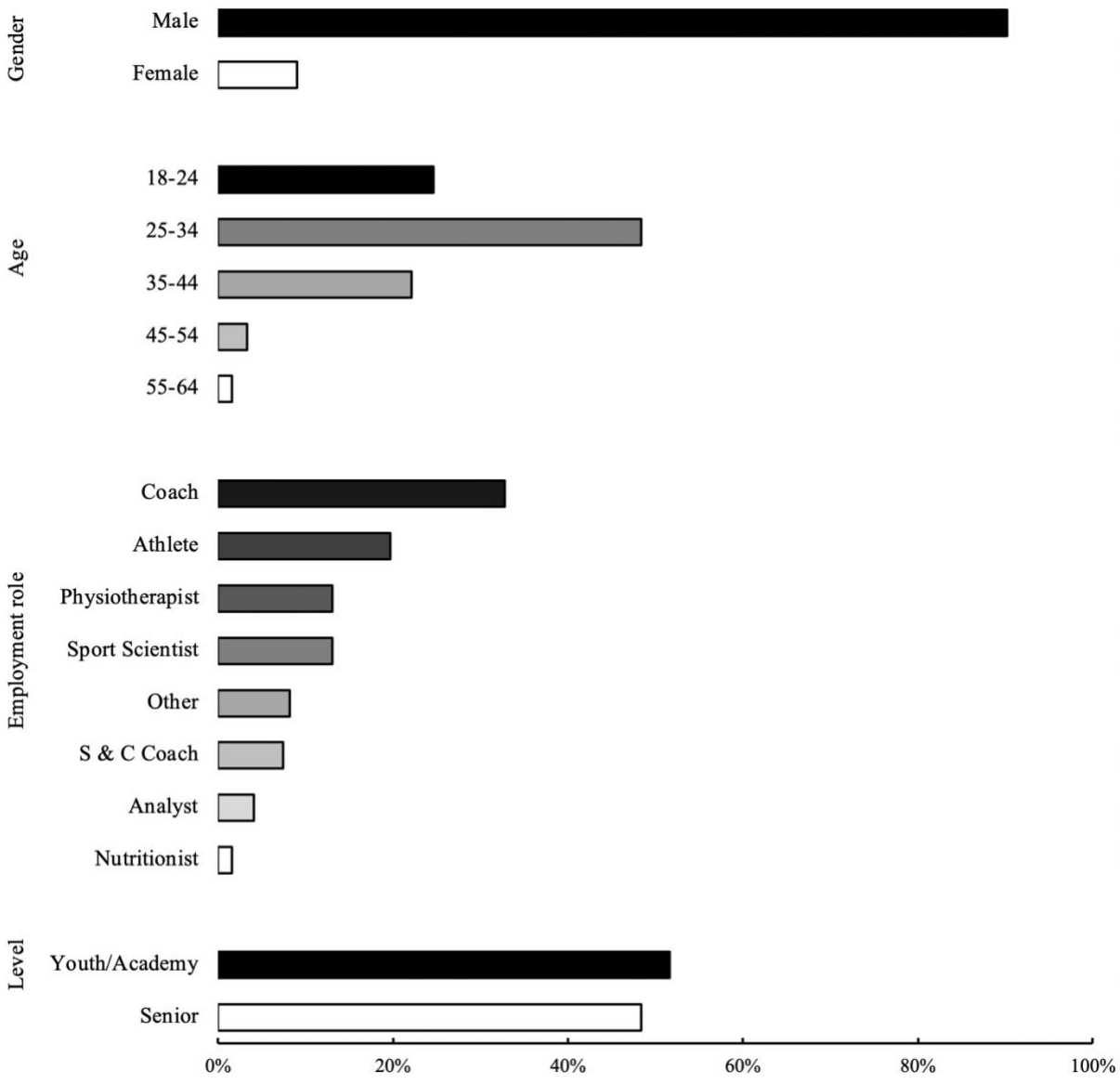


Figure 5.1. Demographics of key stakeholders.

5.3.2 Utilisation

A small minority of respondents reported that they have used genetic testing to aid performance (9%) and/or mitigate injury risk (10%). Similarly, a slightly larger proportion of respondents recounted that an organisation they were employed at has used genetic testing to aid performance (15%) and/or mitigate injury risk (12%). However, the large majority of respondents suggested that they would consider utilising genetic testing in the future for both aiding performance (83%) and/or mitigating injury risk (84%) (see

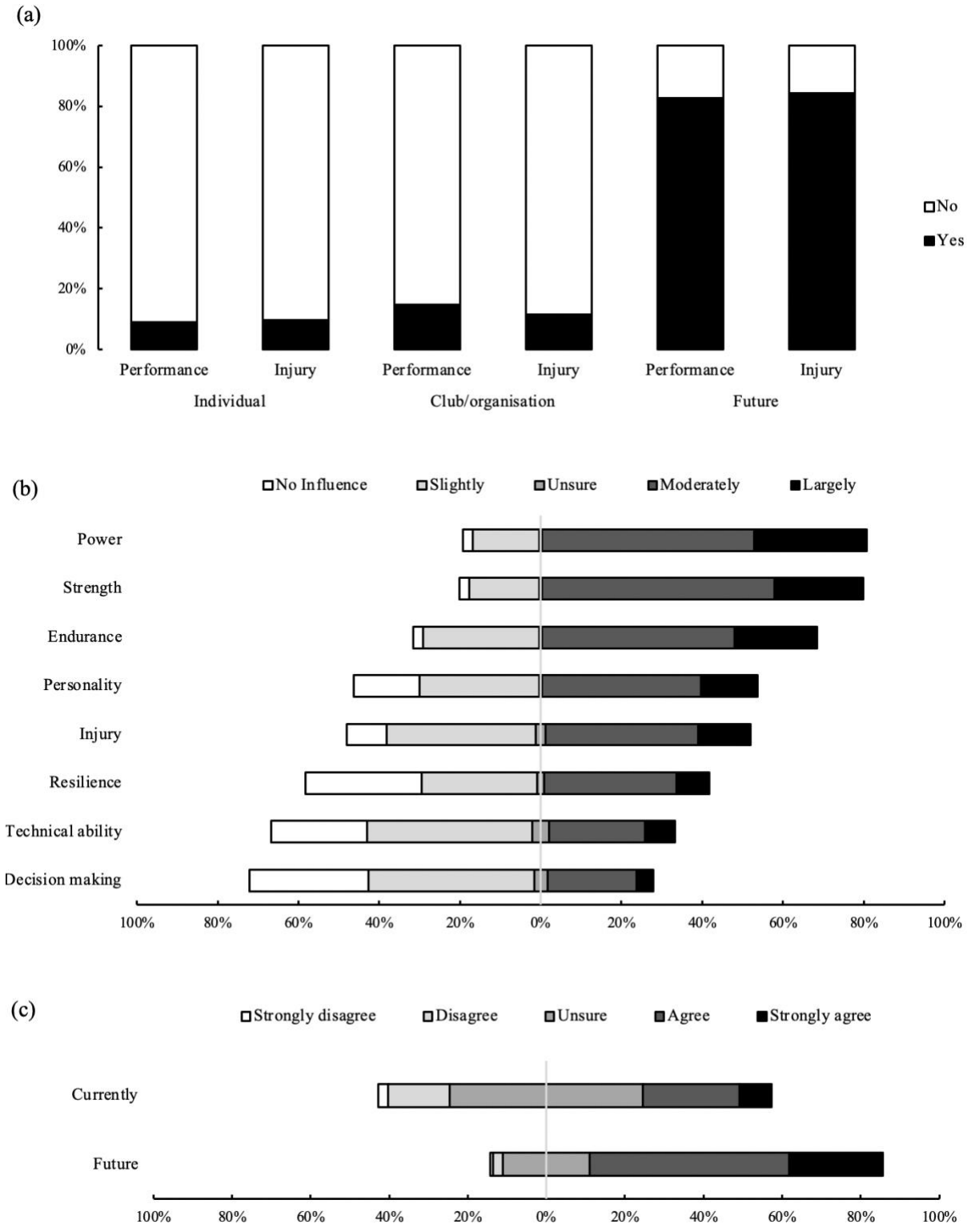


Figure 5.2a). Explanations for selections were provided by some respondents. For example:

"Pretty uncertain how will genetic tests give additional knowledge to what current performance/medical tests can give. The relationship between genetics, epigenetics, and environmental factors is too complex and poorly understood at the moment. It seems they

are deeply interrelated and probably best understood on whole system level. However, maybe advances in our knowledge may make me change my opinion" (Coach, aged 25-34).

5.3.3 Awareness

Overall, power was viewed as the most genetically influenced trait, as it amassed the most 'largely' selections (28%), along with the largest combined 'moderately' and 'largely' selection percentage (81%). In contrast, decision-making was viewed as the least genetically influenced trait, as it amassed the most 'no influence' selections (30%), and similarly the largest combined 'slightly' and 'no influence' selection percentage (71%) (see

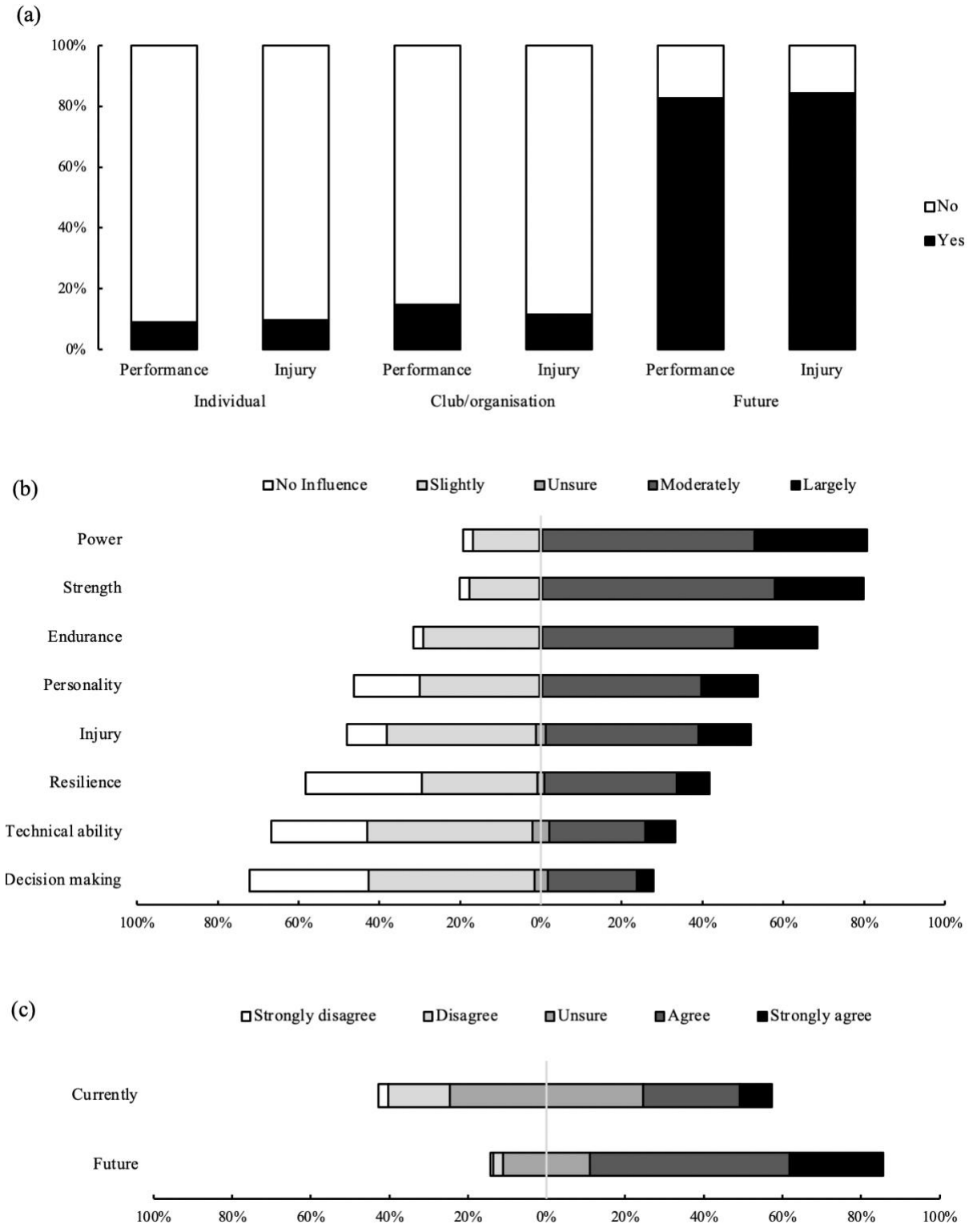


Figure 5.2b). There was a clear pattern in the data showcasing that the majority of key stakeholders perceive genetics to be less influential on psychological-related traits (i.e., decision-making, technical ability, resilience), and more influential on physiological-related

traits (i.e., power, strength, endurance). Explanations for selections were provided by some respondents. For example:

> "Strength and power capabilities from what I have experienced always seem to be the most difficult to make significant and meaningful gains. Some players appear to make greater improvements doing similar or sometimes even less work than others. I typically see this is as a genetic advantage. The other categories in my experience can be more heavily influenced by environmental factors, the athlete's upbringing and training history" (Strength and Conditioning Coach, aged 25-34).

> "Physical attributes can be bettered through training, but the baseline is dependent on genetics. Decision making is affected by experience and knowledge, which can only be learned. I believe nurture rules over nature for personality" (Performance Analyst, aged 25-34).

> "I believe technical ability, decision making, resilience and personality are learnt practises whereas I believe most of the others can be influenced by genetics to some degree" (Athlete, aged 18-24).

5.3.4 Impact

Overall, the opinion of key stakeholders regarding the current utility of genetic testing in sport varied considerably, with 49% unsure, 33% agree/strongly agree, and 18% disagree/strongly disagree. However, the majority agree/strongly agree (75%) that in the future genetic testing will have great utility (see

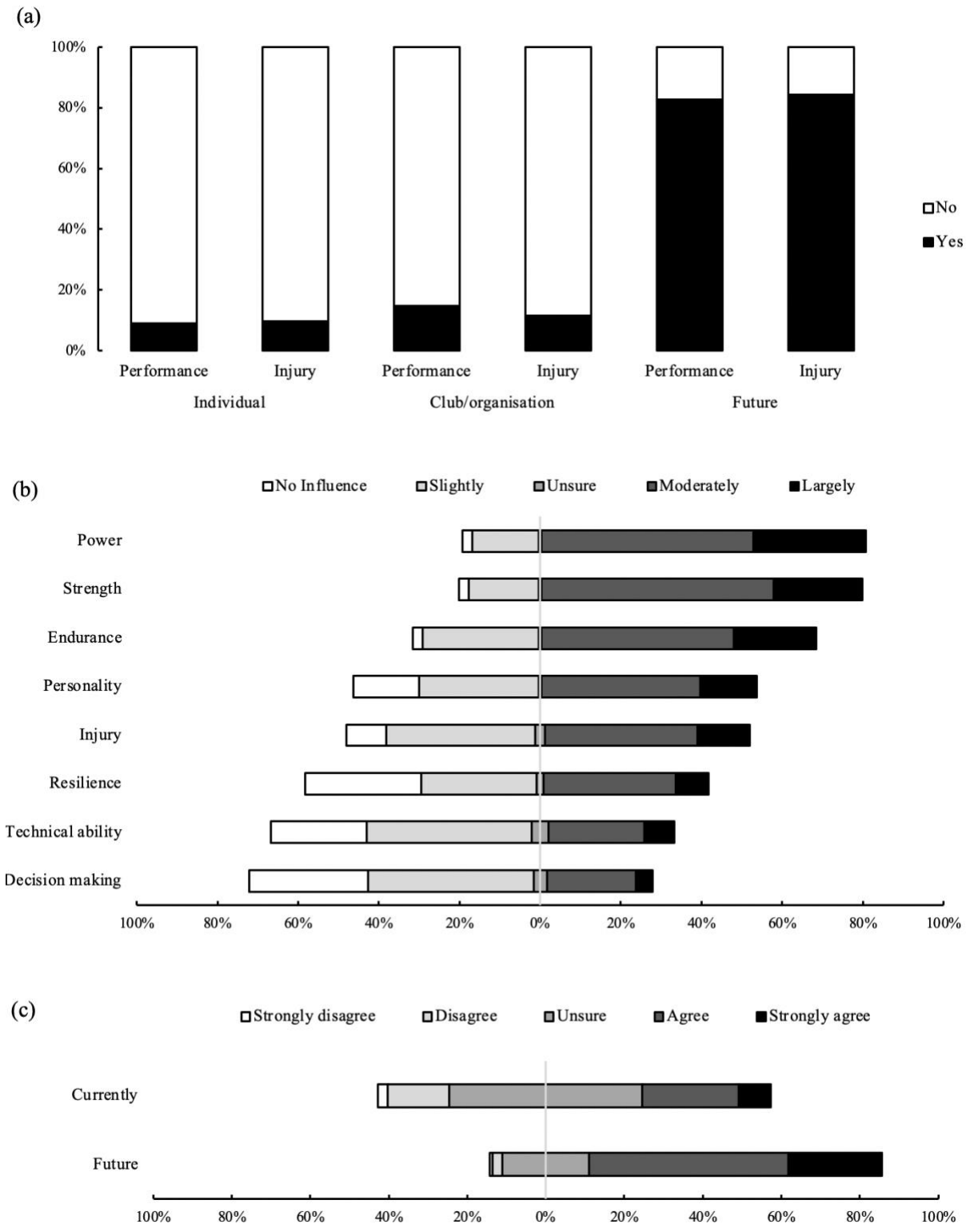


Figure 5.2c). Explanations for selections were provided by some respondents. For example:

"Currently an emerging area of research (relatively speaking) in the field of sport science. Not fully understood in health/medical fields yet, so much more research required before application can be considered" (Nutritionist, aged 35-44).

“Too little is known at the moment regarding the influence of genetics in sport, the work (and sample sizes) are way below anything known in disease” (Sport Scientist, aged 35-44).

“Genetics plays a small role in performance. Data on genetic factors can contribute towards overall analysis on a players' performance” (Coach, aged 18-24).

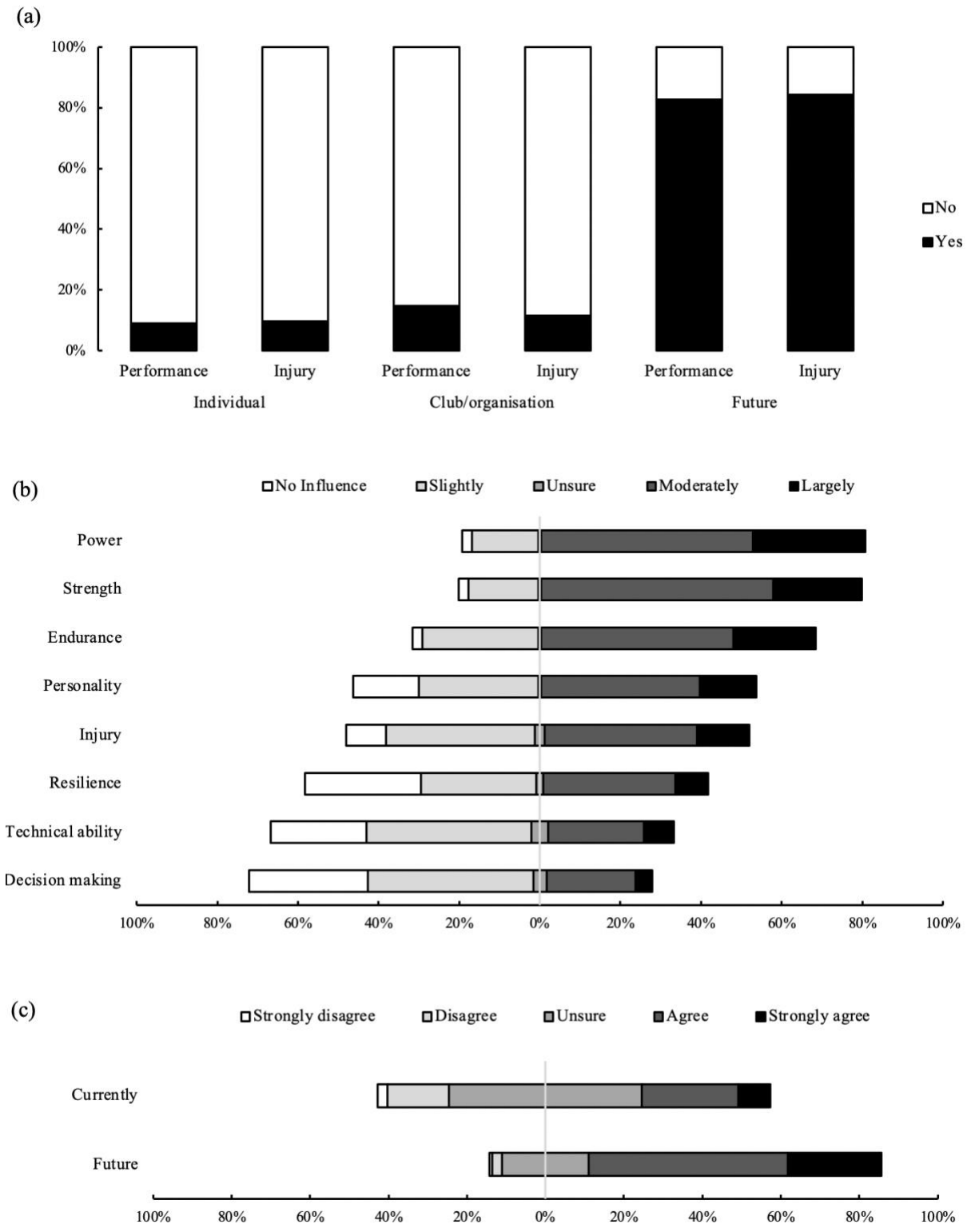


Figure 5.2. Selection percentages of key stakeholders concerning: (a) utilisation; (b) awareness; (c) impact.

5.3.5 Implementation

5.3.5.1 How

Overall, the opinion of key stakeholders regarding the use of genetic testing for talent identification/selection is unclear, as selections were almost evenly distributed: disagree/strongly disagree (36%), agree/strongly agree (35%), and unsure (29%). However, for athlete development/training and injury risk/prevention, the majority of respondents support genetic testing (agree/strongly agree = 70%; 74%) (see **Figure 5.3a**). Explanations for selections were provided by some respondents. For example:

> "I feel it is unethical to select a player based on their genetic potential or ability, they should be identified or selected on merit. When they're in the system genetic testing should then be used to optimise their potential" (Strength and Conditioning Coach, aged 25-34).

> "I think it would definitely help in talent identification and picking players for certain sports and positions in respective sports. This is already done especially in America where the majority of sports science research comes due to the level of funding" (Physiotherapist, aged 18-24).

> "Research (that I'm aware of/read) the focus is on injury risk/prevention and would help provide an insight (by no means the only factor) into a risk that could be mitigated by adapted training, nutrition etc" (Nutritionist, aged 35-44).

5.3.5.2 Barriers

Overall, cost (64%) and knowledge/inability to interpret results (51%), amassed the most 'largely' and only selection majorities. Whereas, the largest combined 'moderately' and 'largely' selection percentages were: (a) cost (83%), (b) knowledge/inability to interpret results (81%), (c) time (68%), and (d) ethical issues (53%) (see **Figure 5.3b**). Explanations for selections were provided by some respondents. For example:

"For bigger clubs who have the money and time for long term results it would be beneficial, but for smaller clubs that need short term success to even just stay afloat then it wouldn't be an option. A lot of management are stuck in tradition and recruit players on their/coaches opinion and don't rely on data to tell them" (Performance Analyst, aged 25-34).

"Knowledge and inability to interpret results have the biggest impact. If we are sure they will help, someone will do it to gain an advantage. I am also not sure that any future knowledge and ability to interpret results will help. There is a probability they just present limited, non-contextual information" (Coach, aged 25-34).

5.3.5.3 Payment

Overall, the majority of respondents expected the pay between £1-100 for a genetic test (71%) and education/training (76%), with the largest selections being £76-100 (25%) and £26-50 (29%), respectively. Whereas, the majority of respondents expected to pay between £26-100 for a consultancy to analyse and interpret results (67%) and genetic counselling to explain results (69%), with the largest selections being £76-100 (26%) and £26-75 (48%), respectively.

5.3.5.4 Access

Regarding who should be allowed access to an athletes genetic data, the respondents selected the following: athlete (24%), scientific staff (19%), parents (14%), anyone athlete consents (14%), sports club (14%), genetic testing company (11%), national government (2%), and sports league/association (2%). Explanations for selections were provided by some respondents. For example:

"I think it should be whoever the athletes consent and the club doctors. The information should belong to the athlete and should be subject to data protection" (Performance Analyst, aged 25-34).

"I think the athlete and the professional club should have access as long as the athlete is an adult i.e. 16 and above" (Coach, aged 35-44).

"Scientific staff < Medical staff. Perhaps player or guardian would need to sign waiver agreeing to results being accessible to club medical staff at the very least" (Sport Scientist, aged 25-34).

5.3.5.5 Influence

Regarding to what extent specific information sources influenced the respondent's opinions, scientific journal articles/conferences amassed the greatest number of 'mostly' selections (22%), along with the largest combined 'mostly' and 'largely' selection percentage (39%). This was followed by professional sport organisations (24%) and word of mouth (22%). In contrast, celebrity endorsement amassed the greatest number of 'no influence' selections (83%), along with the largest combined 'no influence' and 'slightly' selection percentage (94%). This was followed by television/press advertisement (83%), social media (78%), and online videos (76%) (see **Figure 5.3c**).

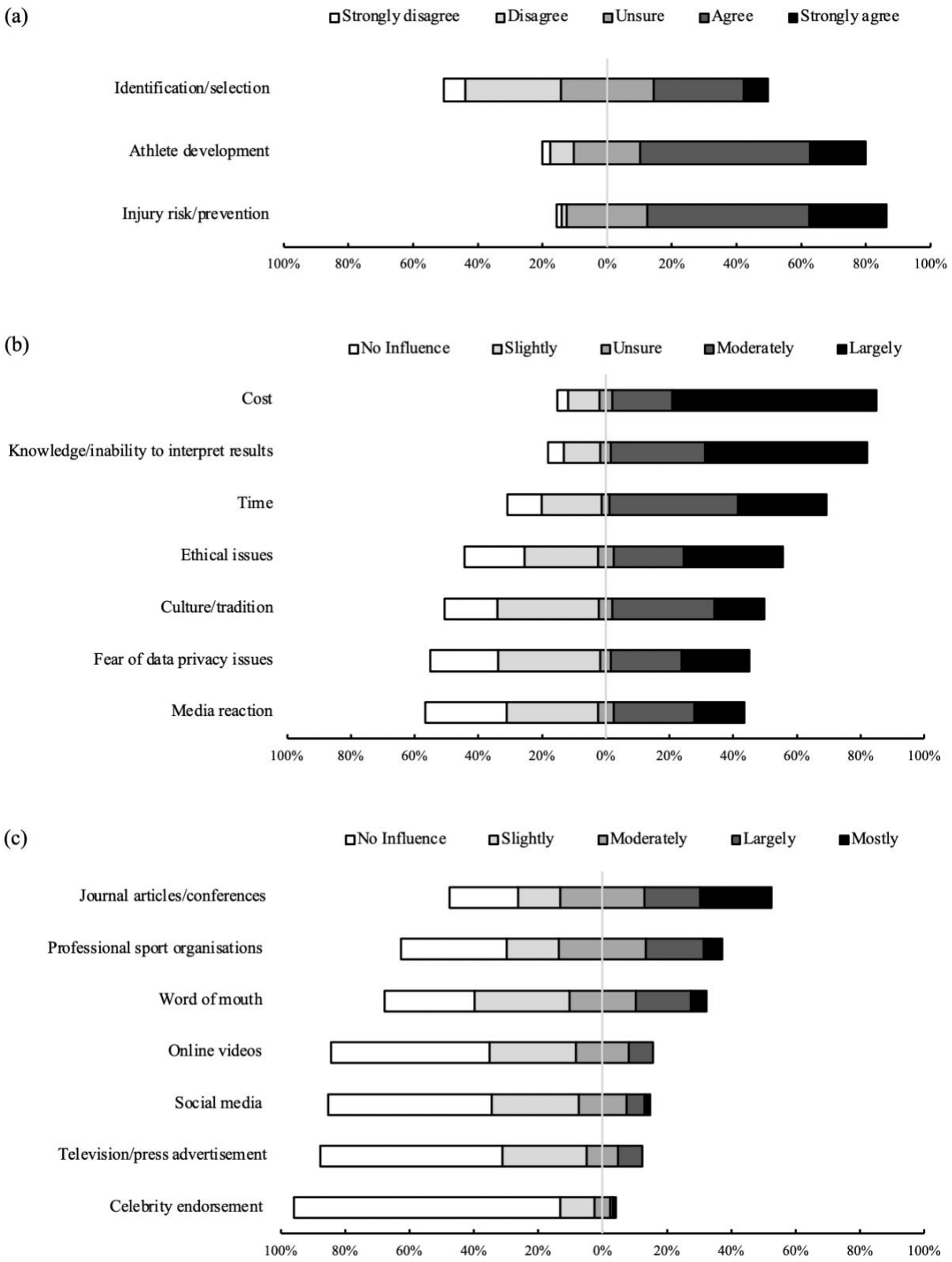


Figure 5.3. Selection percentages of key stakeholders concerning implementation. (a) how; (b) barriers; (c) influence.

5.3.6 Education

The overwhelming majority of respondents indicated that their own individual knowledge (89%), along with their colleague's (79%) of genetic research/testing is insufficient. In addition, the overwhelming majority (91%) believed genetic research is not communicated effectively to key stakeholders. Furthermore, the overwhelming majority indicated that they would be interested in learning more about genetic research in sport (89%), a large number believed that educational workshops are required at their organisation (71%), and a minority believed that the regular opportunity to speak to a genetic specialist or addition of a genetic consultant was required at their organisation (49%). Explanations for selections were provided by some respondents. For example:

> "Specialist subject area – although I aim to actively read relevant genetic research by no means do I consider myself qualified enough in this area, would require some consultation with appropriate professionals" (Nutritionist, aged 35-44).

> "Have BSc in Molecular Biology, but do not feel comfortable to make such decisions" (Coach, aged 25-34).

5.4 Discussion

The purpose of this study was to assess the practical application of genetic testing in professional football by providing an insight into the perspectives of key stakeholders. To the authors' knowledge this is the first study to quantify the use and opinions of genetic testing from a wide range of key stakeholders specifically in football.

Key findings suggest that, although genetic testing is utilised within professional football, it remains relatively rare. Interestingly, genetic testing may currently be used by organisations more frequently compared to four years ago. For example, Varley and colleagues (2018) reported that only 2% and 5% of their multi-sport respondents were aware of their

organisations having ever utilised genetic testing for performance or injury, respectively. In comparison, our current study revealed 15% and 12% of respondents were aware that their organisation had utilised genetic testing for performance and injury, respectively. This coincides with the recent findings of Pickering and Kiely (2021), as 11% of their respondents reported that their organisations have utilised genetic testing. This cumulative 10% increase could be due to a number of possible factors. First, this may represent an increased interest in genetic testing in sport, corresponding with the increase in sport genomic research in recent years (Tanisawa et al., 2020). More specifically in football, McAuley and colleagues (2021a) showed that 55% of genetic association studies involving football players have been published within the last four years. Second, this increased proportion could also illustrate a more accurate representation of the true, and larger, prevalence of genetic testing within professional football. This would not be surprising, as the superior financial situation of football compared to most other sports, especially in Europe (Nauright, 2004), would allow key stakeholders in football to afford a wider spectrum of potential performance measuring metrics. Finally, this increase could just simply be the result of genetic testing becoming more economically viable (Webborn et al., 2015). For instance, the cost of whole-genome sequencing now costs less than $1,000 compared to over $10,000 in 2010 (National Human Genome Research Institute, 2020). Furthermore, the recent surge of DTC companies has inevitably made genetic testing more accessible. Overall, it is most likely a combination of these factors. This is due to significant advances in genomic technologies coupled with rapid reductions in cost outpacing Moore's law (Goodlin et al., 2015a).

If key stakeholders in professional football were to use genetic testing, this study suggests they would mainly do so for physiological-related traits. For instance, the respondents in this cohort believed that genetics have a moderate to large influence on physiological traits, whereas they believed they have a very minor influence on psychological traits. This may be

explained by the paucity of genetic research on psychological traits in football. Indeed, a recent systematic review showcased that out of 80 genetic association studies involving football players, only three studies investigated psychological traits compared to 20 on physiological traits (McAuley, Hughes, et al., 2021a). As such, in football, there is currently a limited psychogenetic evidence base available to key stakeholders to form their opinions on genetic associations with psychological traits. However, it is important to consider that contemporary genetic studies in sport indicate that all traits, irrespective of physiological or psychological foundation, are moderately to highly hereditary (Georgiades et al., 2017). Indeed, the most comprehensive heritability meta-analysis to date (including ~14 million twin pairs and 17,804 human traits) reported a weighted heritability estimate across all human traits of 49% (Polderman et al., 2015). Although, as genetic testing aims to reveal which variants (previously associated with specific traits) an individual possesses, it should be noted that very few psychogenetic variants have been identified and validated in sport (Valeeva et al., 2019). In light of this information, perhaps the perspectives of the key stakeholders should not be surprising regarding the genetic influence on psychological traits.

This study suggests that there is limited knowledge of genetic research and/or testing amongst key stakeholders in professional football. For instance, almost half of the stakeholders (49%) in this study reported that they are unsure of genetic testing's present utility. Moreover, the majority (81%) reported their inability to interpret results as a significant barrier to implementing genetic testing in football. This is exemplified by most of the stakeholders (84%), who believed that no employee at their organisation has sufficient knowledge of genetic research and testing. This may be due to the education methods accessible to stakeholders at football organisations, as an overwhelming majority of stakeholders (91%) indicated that genetic research is not currently communicated effectively with coaches, practitioners, and players. This is potentially problematic, especially when it is anticipated that genetic testing

will become increasingly common, which is reinforced by most stakeholders (75%) who believed genetic testing will have great utility in the future. For instance, it has been reported that China are implementing genetic testing as part of their athlete selection ahead of the 2022 Winter Olympic Games (Haff, 2019). Indeed, a number of football organisations, such as FC Barcelona (Miller, 2016) and the Egyptian Football Association (Holmes, 2018), have recently begun to utilise genetic information for training optimisation and injury prevention. Thus, more effective education may be needed in football. This was also indicated by a large proportion of stakeholders (71%) in this current study, who reported that educational workshops are required at their organisations. It could be suggested, the implementation of an education programme on genetics may be well received in football, as nearly all of the stakeholders (91%) expressed a desire to learn more about genetic research and the validity of genetic testing. As such, future research is required to explore the most efficient and effective approaches to provide evidence-based information on genetic research and practical application in professional football.

Education may be particularly important for key stakeholders in professional football when considering uses of genetic testing. Specifically, although the majority of stakeholders (72%) believed genetic testing should be used for athlete development and/or injury risk, over a third (35%) believed genetic testing should also be utilised for talent identification and selection purposes. This may be considered problematic, as the use of genetic testing for talent identification and selection in sport is currently considered immoral, unethical, and unlikely to give useful information (Patel & Varley, 2019; Vlahovich et al., 2017; Webborn et al., 2015). Indeed, using genetic testing for these purposes could impede child development and constrain their right to an open future (Camporesi & McNamee, 2016). It is also important to consider that genetic research into athlete status has yielded few genetic variants that have been adequately replicated in independent cohorts (Ahmetov et al., 2016). This is due to their small effect sizes and consequent lack of explained inter-individual variance between athletes and

controls (Bouchard, 2015). For example, a recent review meta-analysed the most studied genetic variants associated with athlete status in football (i.e., *ACTN3* R577X and *ACE* I/D), which reported only modest allelic odds ratios of 1.18-1.35 (McAuley, Hughes, et al., 2021b). In the future, it is also highly unlikely any genetic variant, or polygenic profile, will ever have the specificity/sensitivity to solely predict future sporting prowess (Pickering et al., 2019; Williams & Folland, 2008). Indeed, currently only 24.6% of height, which has an estimated heritability of 80%, has been explained by genome-wide significant polymorphisms (Yengo et al., 2018). This is underpinned by the complexity of gene-gene and gene-environment interactions, as well as the multifactorial and dynamic nature of athlete development (Baker et al., 2019; McAuley, Baker, et al., 2021; Tucker & Collins, 2012; Yan et al., 2016). Therefore, as high-performance is not an isolated, independent, or static trait, expertise in the sporting domain may never be fully quantifiable or predicted accurately via any performance measuring metric, including genetic information (Elhert et al., 2013; Mattsson et al., 2016). As a result, key stakeholders in professional football are recommended to act with caution when utilising genetic testing for these purposes, whilst researchers are encouraged to design, implement, and evaluate methods of education.

5.4.1 Limitations

Although this study has amassed the opinions from coaches, practitioners, and players across a wide-range of competitive playing levels in youth and senior football, it should be acknowledged that it is not without its limitations. Specifically, the sample size was relatively small, the majority of respondents were male, and the views expressed are not representative of the entire football ecosystem (e.g., no responses from medical doctors). Previous research has also demonstrated that there are gender, age, educational, and social/cultural related differences in the acceptance of genetic testing (Aro et al., 1997), which may have influenced the results based on our cohort demographics. Moreover, economic differences between

countries and organisations may have influenced the results as the authors' pre-existing contacts were predominately U.K based. Since the survey was anonymous, sampling bias may also have skewed responses as it is unknown how many key stakeholders were from the same organisation. Furthermore, as the survey was circulated on social media, this may have biased the sample as the authors' followers may be interested in genetic testing. Finally, the questions within the survey were not validated, but did comprise of similar questions to Varley et al. (2018) and Pickering and Kiely (2021) to allow comparisons. Moving forward, it is important that future research considers the use of a validated survey to capture more consistent datasets. Nevertheless, the collected responses provide a useful preliminary assessment of the existing knowledge and application of genetic testing in professional football from a diverse cohort of key stakeholders.

5.5 Conclusion

This study suggests that genetic testing is rarely used within professional football. However, key stakeholders anticipate that it will be utilised more in the future. Given the perceived lack of knowledge and education, implementation of education programmes may prove valuable in improving key stakeholders' knowledge and the practical application of genetic testing in professional football. Further studies using larger cohorts with more representation from different stakeholders (e.g., medical doctors) and greater diversity within the roles represented (e.g., females) are encouraged in order to determine and validate the perspectives of key stakeholders. In addition, future studies should explore what methods are most effective in providing key stakeholders with evidenced-based information on genetic research and testing. Ultimately, it will be essential that future research critically examines the practical application of genetic testing within professional football.

6 Genetic associations with technical capabilities in English academy football players: A preliminary study

McAuley, A. B. T., Hughes, D. C., Tsaprouni, L. G., Varley, I., Suraci, B., Baker, J., Herbert, A. J., & Kelly, A. L. (2022). Genetic associations with technical capabilities in English academy football players: a preliminary study. *The Journal of sports medicine and physical fitness*. Advance online publication. https://doi.org/10.23736/S0022-4707.22.13945-9

Abstract

Technical capabilities have significant discriminative and prognostic power in youth football. Although, many factors influence technical performance, no research has explored the genetic contribution. As such, the purpose of this study was to examine the association of several single nucleotide polymorphisms (SNPs) with technical assessments in youth football players. Fifty-three male under-13 to under-18 outfield football players from two Category 3 English academies were genotyped for eight SNPs. Objective (n = 26) and subjective (n = 27) technical performance scores in dribbling, passing, and shooting were collated from two independent cohorts. Simple linear regression was used to analyse individual SNP associations each variable, whereas both unweighted and weighted total genotype scores (TGSs; TWGSs) were computed to measure the combined influence of all SNPs. In isolation, the *ADBR2* (rs1042714) C allele, *BDNF* (rs6265) C/C genotype, *DBH* (rs1611115) C/C genotype, and *DRD1* (rs4532) C allele were associated with superior (8-10%) objective dribbling and/or shooting performance. The TGSs and/or TWGSs were significantly correlated with each technical assessment (except subjective passing), explaining up to 36% and 40% of the variance in the objective and subjective assessments, respectively. As such, this study suggests inter-

individual genetic variation may influence the technical capabilities of youth football players and proposes several candidate SNPs that warrant further investigation.

6.1 Introduction

In football (soccer), technical skills (e.g., dribbling, passing, and shooting) are some of the most frequently performed actions during match-play (Williams et al., 2020). There has been a significant increase in the frequency of technical skills completed within national and international competitions over time. For instance, an increased number of crosses, dribbles, and passes were attempted from the 1991/92 to the 1997/98 season in the top tier of English football (Williams et al., 1999). Similarly, from the 2006/07 to the 2012/13 season in the English Premier League there was an increased incidence of dribbling (~20%) and passing (~30%; Barnes et al., 2014). In the context of international competition, ball speeds (~15%) and passing rates (~35%) increased significantly between 1966 and 2010 in World Cup final matches (Wallace & Norton, 2014).

The continued rise in the number of technical actions performed during match-play in football may be driven by their association with the relative success of the team. Indeed, Rampinini et al. (2009) reported that more dribbling, passing, and shooting were associated with a higher league position (1^{st} to 5^{th} vs. 16^{th} to 20^{th}) in the Italian Serie A. Moreover, technical actions such as passing and shooting frequency were related to a team's league position (1^{st} to 6^{th} vs. 7^{th} to 18^{th}) in the Greek Superlegue first division (Gómez et al., 2018). Similar findings have also been reported in youth football, whereby teams that attempted a greater number of shots attained a higher placing (1^{st} to 4^{th}) at the under-17 Al Kass International Cup (Varley et al., 2017).

Technical capabilities also have significant discriminative power with regards to individual playing levels and perceived potential within youth football (Koopmann et al., 2020). For instance, Vaeyens et al. (2006) showcased that a technical testing battery (ball juggling, lob pass, shooting accuracy, and slalom dribble) could distinguish under-13 to under-16 ability groups (i.e., elite, sub-elite, and non-elite). More recently, Kelly et al. (2020) reported

that ‘higher-potential’ under-12 to under-16 players in the same academy performed significantly better at the same four technical tests than ‘lower-potentials’. In addition, Keller et al. (2016) revealed that the Loughborough Short Passing Test, long passing test, shooting test, and speed dribbling test were able to differentiate under-18 youth football players, with the highest playing level achieving the highest scores.

From a talent identification perspective, perhaps the most important factor is the significant prognostic value of technical skill possessed at younger ages on career progression and attainment (Murr et al., 2018). As an example, Huijgen et al. (2009) reported that professional players outperformed amateur players in dribbling during ages 14 to 18 years in the Netherlands. Similarly, Forsman et al. (2016) revealed that professional players outperformed non-professional players in dribbling and passing at age 15 years in Finland. In Germany, professionals outperformed both semi-professionals and non-professionals in ball control, dribbling, and shooting at the under-12 to under-15 age groups (Höner et al., 2017; Leyhr et al., 2018).

Given the evidence supporting the influence of technical skills on a football match as well as their discriminative and prognostic power, it is not surprising coaches and scouts view technical skill as an integral part of evaluating a player’s current performance and future potential (Roberts et al., 2019). However, technical capabilities are the result of a complex and multifactorial relationship between several factors. Indeed, age, maturation, lean body mass, hours of practice, type of practice, and playing position have all been associated with influencing technical skill in some capacity (Sarmento et al., 2018). Although, a potentially important factor that is currently under-researched throughout talent identification and development in football is the contribution of genetics (McAuley, Hughes, et al., 2021a).

More specifically, inter-individual genetic variation is responsible for differences in all observable human traits to some extent (Polderman et al., 2015). Heritability studies have

reported relatively high estimates for the influence of genetics on traits theoretically associated with technical capability, such as motor control and motor learning (~70%) (Fox et al., 1996; Missitzi et al., 2004, 2011, 2013, 2018; Pellicciari et al., 2009). A number of genetic polymorphisms have also already been identified that influence several of these traits, particularly polymorphisms in genes that encode for dopamine receptors and degradation enzymes within the dopaminergic system (Kitazawa et al., 2021; Tunbridge et al., 2019). These polymorphisms are consistently associated with traits such as motor control and motor learning due to the role of dopamine in several brain processes, including learning, movement, plasticity, and reward (Pearson-Fuhrhop et al., 2013). Therefore, as a by-product, these genetic polymorphisms may also be related to technical ability in football.

The extent to which genetic variants may contribute to technical capabilities within the context of football is very limited. In Australian Rules Football (AFL), Jacob et al. (2018) found associations between several single nucleotide polymorphisms (SNPs) in the *ACE*, *ADRB2*, *ADRB3*, *BDNF*, *COMT*, and *DRD2* genes and the Nathan Buckley kicking skill assessment. Although SNPs appear to impact technical skills in AFL, it is currently unclear to what extent inter-individual genetic variation influences technical skill specifically in football. Moreover, the genetic polymorphisms that are ultimately responsible for any association are yet to be identified. Therefore, this preliminary study examined associations between genetic polymorphisms and technical assessments in youth football players.

6.2 Methods

6.2.1 Participants

Fifty-three male under-13 to under-18 (aged 16.28 ± 1.27 years) outfield football players from two Category 3 English academies were examined. Objective technical performance scores (n = 26) were collected from one academy and subjective technical coach ratings (n = 27) were collected from the other academy, in adherence with each club's standard player assessment

protocols. Informed assent from all players, consent from parents/guardians, and gatekeeper consent from each academy was collected prior to the commencement of the study. All experimental procedures were conducted in accordance with the guidelines in the Declaration of Helsinki and ethical approval was granted by Birmingham City University via the Health, Education, and Life Sciences Academic Ethics Committee. This study was conducted in accordance with the recommendations for reporting the results of genetic association studies defined by the STrengthening the REporting of Genetic Association studies (STREGA) Statement (Little et al., 2009).

6.2.2 Objective data

Objective data on three football-specific technical tests were collected: (a) slalom dribble, (b) lob pass, and (c) shooting accuracy. The slalom dribble test required players to control the ball through nine cones (2 m apart) from the start to the end line and return. Each player completed two trials and the quickest was recorded for analysis, measured via timing gates (Brower TC Timing System, Draper, Utah, USA). In the lob pass test, players had ten attempts (five with each foot) to kick and elevate the football from a distance of 20 m into a target area divided into three concentric circles. Each circle was different in diameter (3 m, 6 m, and 9.15 m) and each attempt received points (3, 2, and 1, respectively) depending on the circle in which the ball originally landed. In the shooting accuracy test, players had ten attempts (five with each foot) to kick the ball at a 16 m wide goal target from a shooting distance of 20 m central to the goal. The goal was divided into five parallel zones: centre (2 m wide), two areas on each side of the centre (3 m wide), and two areas 4 m wide at each extreme (4 m wide) that award different points (3, 2, and 1, respectively). All tests have been previously utilised in football research as valid indicators of technical skill in youth football populations (Kelly et al., 2020; Vaeyens et al., 2006).

6.2.3 Subjective data

Subjective data on the technical skills of dribbling, passing, and shooting were collected. Two coaches from each age group (who possessed the minimum of a UEFA B licence) used a 5-point Likert scale to rate the technical abilities of each player from their respective age group: 1 = significant weakness; 2 = requires attention; 3 = competent; 4 = accomplished; and 5 = excellent. Each coach independently completed their technical ratings. The average rating on each technical skill from all coaches was then recorded for analysis. The coach-based subjective ratings used in this study have been utilised previously by researchers in youth football and demonstrate good reliability and validity (Ali, 2011; Dugdale et al., 2020; Henry et al., 2018).

6.2.4 Genetic procedures

6.2.4.1 Genotyping

Saliva was collected from players via sterile, self-administered buccal swabs, following a minimum of 30 minutes since food or drink ingestion. Saliva samples were safely stored at room temperature and within 36 hours were sent to AKESOgen, Inc. (Peachtree Corners, GA, USA) for DNA extraction. Using Qiagen chemistry, DNA was extracted on an automated Kingfisher FLEX instrument (Thermo Fisher Scientific, Waltham, MA, US). The manufacturer's recommended guidelines and procedures were followed throughout. To measure the extracted DNA's quality and quantity, PicoGreen and Nanodrop measurements were taken. Input to the custom testing array occurs at 200 ng in 20 μL. Amplification, fragmentation, and resuspension were performed using Biomek FXP. GeneTitan instrumentation (Thermo Fisher Scientific, Waltham, MA, US) was used to stain and scan the arrays, with hybridisation performed in a Binder oven at 48 degrees for 24 hours, following the Affymetrix Axiom high throughput 2.0 protocol. Data analysis was then performed using a raw

CEL file data input into the Affymetrix Axiom Analysis Suite (Affymetrix, Santa Clara, CA, US). Procedures are in accordance with Pickering et al. (2019).

6.2.4.2 Polymorphism selection

To identify polymorphisms potentially associated with technical skills in football, empirical research, review articles, book chapters, and the GWAS catalog (https://www.ebi.ac.uk/gwas/) were examined. Priority was given to GWAS results that were replicated in independent cohorts, followed by candidate gene studies with large homogenous sample sizes which produced similar associations in more than one study. The polymorphisms that were finally included were dependent on the coverage of the microarray and quality control procedures. After an extensive search of the literature, the following polymorphisms were selected based on their proposed biological function and relevant associations in previous studies: Adrenoceptor beta 2 (*ADRB2*; rs1042714), Angiotensin I converting enzyme (*ACE*; rs4341), Brain derived neurotrophic factor (*BDNF*; rs6265), Catechol-O-methyltransferase (*COMT*; rs4680), Dopamine beta-hydroxylase (*DBH*; rs1611115), Dopamine receptor D1 (*DRD1*; rs4532), Dopamine receptor D2 (*DRD2*; rs1076560), and Dopamine receptor D3 (*DRD3*; rs6280) (see **Table 6.1**). These gene names and symbols are in accordance with those officially approved by the Human Gene Nomenclature Committee (HGNC; https://www.genenames.org). Standard genomic quality control (QC) procedures and thresholds were applied when selecting polymorphisms: SNP call rate (>95), sample call rate (>95), fisher's linear discriminant (>3.6), and minor allele frequency (>0.05).

Table 6.1. Technical gene and single nucleotide polymorphism (SNP) information.

Gene	Symbol	Chr	Function	SNP	Consequence	Associations	MAF
Adrenoceptor beta 2	*ADRB2*	5q32	Expressed in the brain as part of the catecholamine system and acts as a receptor for adrenaline, noradrenaline, and dopamine.	rs1042714	Missense variant G > C (Glu > Gln)	Receptor activity and sensitivity, neurodevelopmental disorders, white matter and cognitive functions.	G = 0.41
Angiotensin I converting enzyme	*ACE*	17q23.3	Catalyses the degradation of the inactive decapeptide angiotensin I, and generates the physiologically active peptide, angiotensin II.	rs4341	Intron variant C > G (Insertion > Deletion)	Serum and tissue activity, blood pressure, percentage of muscle fibre types, strength and muscle volume.	C = 0.43
Brain derived neurotrophic factor	*BDNF*	11p14.1	Encodes for a neurotrophin essential for neuronal development and survival, synaptic plasticity, and cognitive function.	rs6265	Missense variant C > T (Val > Met)	Motor cortex plasticity, motor learning, neuromuscular activation, horizontal and vertical power.	T = 0.20
Catechol-O-methyltransferase	*COMT*	22q11.21	Catalyses the degradation of catecholamines, including neurotransmitters dopamine, epinephrine, and norepinephrine.	rs4680	Missense variant G > A (Val > Met)	Enzymatic activity, dopaminergic tone, F-Dopa metabolism, motor learning, working memory, executive functioning and visuospatial tasks.	A = 0.50
Dopamine beta-hydroxylase	*DBH*	9q34.2	Catalyses the oxidative hydroxylation of dopamine to norepinephrine modulating executive and motor function.	rs1611115	2KB Upstream variant C > T	Enzymatic activity, metabolism of dopamine and norepinephrine, as well as neurodevelopmental disorders.	T = 0.21
Dopamine receptor D1	*DRD1*	5q35.2	Encodes the D1 subtype of the dopamine receptor, which stimulates adenylyl cyclase and activates cyclic AMP-dependent protein kinases, regulating neuronal growth, development, and behaviour.	rs4532	5 Prime UTR variant C > T	Dopaminergic tone, dopamine neurotransmission, motor learning, neurodevelopmental and neurodegenerative disorders.	C = 0.40
Dopamine receptor D2	*DRD2*	11q23.2	Encodes the D2 subtype of the dopamine receptor, which inhibits adenylyl cyclase activity and regulates neuronal growth, development, and behaviour.	rs1076560	Intron variant C > A	Dopaminergic tone, dopamine neurotransmission, motor learning, neurodevelopmental and neurodegenerative disorders.	A = 0.15
Dopamine receptor D3	*DRD3*	3q13.31	Encodes the D3 subtype of the dopamine receptor, which inhibits adenylyl cyclase activity and is associated with cognitive and emotional functions.	rs6280	Missense variant C > T (Gly > Ser)	Dopaminergic tone, dopamine neurotransmission, motor learning, neurodevelopmental and neurodegenerative disorders.	C = 0.33

Note. Chr = Chromosome location; MAF = Minor allele frequency (according to European population; 1000 Genomes Project Consortium, 2015).

6.2.4.3 Total genotype score

A total genotype score (TGS) was calculated to assess the combined influence of the included SNPs on each dependent variable. Since Williams and Folland (2008) first proposed and implemented the TGS, the mathematical algorithm has undergone various modifications to try and improve its accuracy. For instance, one approach has included incorporating a mathematical weight for each SNP based on its partial influence in a regression model (Massidda et al., 2014; Varillas Delgado et al., 2020). Both unweighted and weighted TGS approaches have demonstrated sufficient discriminatory power (Charlier et al., 2017). As such, both an unweighted and weighted TGS were calculated and implemented in this study (referred to herein as TGS and TWGS, respectively).

To generate both the TGS and TWGS, each genotype of a respective SNP initially received a score between 0-2 based on the observed genotype associations with a dependent variable. Genotypes of co-dominant models (AA vs. Aa vs. aa) were assigned three scores (i.e., homozygous-associated genotypes received a score of two, the heterozygote received a score of one, and the alternate homozygous genotype received a score of zero), whereas genotypes of dominant (AA vs. Aa-aa) and recessive (AA-Aa vs. aa) models were assigned a score of two (i.e., associated genotype[s]) or zero (i.e., alternate genotype[s]) (Guilherme et al., 2020).

For the TGS, the genotype scores (GS) were then summed and transformed into a 0-100 scale by dividing the total score by the maximum possible score and multiplying by 100.

$$TGS = (combined\text{-}GS / maximum\text{-}GS) * 100$$

For the TWGS, a similar procedure to Varillas Delgado et al. (2020) was used. Each GS was multiplied by the standardised beta coefficients (β) of each SNP following multiple regression with each dependent variable to create weighted genotype scores (WGS). The WGSs were then summed and transformed into a 0-100 scale by dividing the total score by the

maximum possible score and multiplying by 100. Greater values in both models indicate a more advantageous polygenic profile.

$$TWGS = (combined\text{-}WGS / maximum\text{-}WGS) * 100$$

6.2.5 Data analysis

Each SNP was tested for adherence with Hardy-Weinberg equilibrium (HWE) using an exact test via SNPStats (Solé et al., 2006). Linkage disequilibrium (LD) was analysed using LDlink (Machiela & Chanock, 2015) and data from the 1000 Genomes Project European ancestry population (1000 Genomes Project Consortium, 2015). All other data were analysed using Jamovi version 1.8.1 and IBM SPSS version 25. Normality was assessed with the Shapiro-Wilk test and homoscedasticity was assessed using Levene's test. Akaike information criterion (AIC) was used to select which genetic model (i.e., co-dominant, dominant, recessive) best fits the data and would be subjected to hypothesis testing. However, if MAF ≤ 0.25, a dominant model was utilised to retain statistical power (Murtagh et al., 2020). Simple linear regression was performed to assess the association of genotype models with each objective and subjective dribbling, passing, and shooting assessments. Age was controlled for in the objective analysis by adding it as a covariate. Multiple regression was used to calculate the standardised beta coefficients (β) of each SNP for the TWGS models. Simple linear regression was then performed to assess the association of each TGS and TWGS with each dependent variable. Pearson's correlation coefficient (r) with thresholds values of ≤0.1 (trivial), >0.1-0.3 (small), >0.3–0.5 (moderate), >0.5–0.7 (large), >0.7–0.9 (very large), and >0.9–1.0 (almost perfect) were used to measure correlation (Hopkins, 2009). The coefficient of determination (R^2) was computed to determine the variance explained by each TGS and TWGS. Statistical significance was set at $p < .05$.

6.3 Results

Genotype and allele distributions of all SNPs were in HWE and all SNPs were in linkage equilibrium. Assumptions of normality and homoscedasticity were not violated. Descriptive statistics for objective and subjective technical variables and genotype frequencies are displayed in **Table 6.2** and **Table 6.3**, respectively.

6.3.1 Objective associations

6.3.1.1 Dribbling

There was a significant association between *DRD1* ($F_{(1, 23)} = 6.51$, $p = .018$), *ADBR2* ($F_{(1, 23)} = 6.32$, $p = .019$), *DBH* ($F_{(1, 23)} = 4.64$, $p = .042$), and dribbling performance (see **Table 6.4**). More specifically, *DRD1* T/T homozygotes were 10% slower than C allele carriers ($B = 1.37$), *ADBR2* G/G homozygotes were 9.5% slower than C allele carriers ($B = 1.46$), and *DBH* C/C homozygotes were 8.2% faster than heterozygotes ($B = -1.27$). There were also significant associations with both the TGS ($F_{(1, 24)} = 7.02$, $p = .014$) and the TWGS ($F_{(1, 24)} = 8.74$, $p = .007$) and dribbling. While the TGS had a moderate negative correlation ($r = -.48$) and explained 23% of the variance, the TWGS had a large negative correlation ($r = -.52$) and explained 27% of the variance (see **Figure 6.1**).

6.3.1.2 Passing

There were no significant associations between any single SNP or the TGS ($F_{(1, 24)} = 1.33$, $p = .260$) and passing performance. However, there was a significant association with the TWGS and passing ($F_{(1, 24)} = 5.23$, $p = .031$), which had a moderate positive correlation ($r = .42$) and explained 18% of the variance (see **Figure 6.2**).

6.3.1.3 Shooting

There was a significant association between *BDNF* ($F_{(1, 23)} = 6.78$, $p = .016$), *DBH* ($F_{(1, 23)} = 4.52$, $p = .044$), and shooting performance. More specifically, *BDNF* C/C homozygotes achieved 9.5% higher scores than T allele carriers ($B = 2.27$) and *DBH* C/C homozygotes

achieved 8.2% higher scores than T allele carriers (B = 1.96). There were also significant associations with the TGS ($F_{(1, 24)}$ = 10.86, p = .003) and the TWGS ($F_{(1, 24)}$ = 13.74, p = .001) and shooting. Both the TGS and TWGS had large positive correlations (r = .56; r = .60) and explained 31% and 36% of the variance, respectively (see **Figure 6.3**).

6.3.2 Subjective associations

6.3.2.1 Dribbling

There were no significant associations between any single SNP and dribbling rating (see **Table 6.5**). However, there were significant associations with the TGS ($F_{(1, 25)}$ = 8.91, p = .006) and the TWGS ($F_{(1, 25)}$ = 16.71, p < .001). More specifically, both the TGS and TWGS had large positive correlations (r = .51; r = .63) and explained 26% and 40% of the variance, respectively (see **Figure 6.4**).

6.3.2.2 Passing

No significant associations were identified between any single SNP, the TGS ($F_{(1, 25)}$ = 2.38, p = .135), or the TWGS ($F_{(1, 25)}$ = 3.08, p = .092) and passing rating (see **Figure 6.5**).

6.3.2.3 Shooting

There were no significant associations between any single SNP and shooting rating. However, there were significant associations with the TGS ($F_{(1, 25)}$ = 7.85, p = .010) and the TWGS ($F_{(1, 25)}$ = 11.97, p = .002). The TGS had a moderate positive correlation (r = .49) and explained 24% of the variance, whereas the TWGS had a large positive correlation (r = .57) and explained 32% of the variance (see **Figure 6.6**).

Table 6.2. Descriptive statistics of objective technical scores.

Gene (SNP)	Genotype = *n* (%)	Dribbling	Passing	Shooting	MAF	HWE
Overall		15.17 ± 1.98	16.50 ± 5.44	24.12 ± 2.34	N/A	N/A
ADBR2 (rs1042714)	G/G = 8 (31)	15.95 ± 1.77	17.25 ± 6.25	22.88 ± 2.10	0.44	1
	G/C = 13 (50)	14.26 ± 1.60	17.23 ± 4.49	24.46 ± 2.37		
	C/C = 5 (19)	16.31 ± 2.40	13.40 ± 6.43	25.20 ± 2.17		
ACE (rs4341)	G/G = 7 (27)	14.85 ± 2.43	16.29 ± 6.13	23.71 ± 3.64	0.46	1
	G/C = 14 (54)	15.16 ± 1.69	16.93 ± 5.15	24.43 ± 1.74		
	C/C = 5 (19)	15.66 ± 2.41	15.60 ± 6.39	23.80 ± 1.92		
BDNF (rs6265)	C/C = 16 (62)	15.48 ± 1.89	16.06 ± 5.01	24.88 ± 1.96	0.21	1
	C/T = 9 (35)	14.38 ± 2.00	18.44 ± 5.20	22.67 ± 2.50		
	T/T = 1 (4)	17.43 ± N/A	6.00 ± N/A	25.00 ± N/A		
COMT (rs4680)	A/A = 6 (23)	14.88 ± 1.76	18.50 ± 6.12	24.50 ± 3.73	0.48	0.69
	A/G = 15 (58)	15.48 ± 1.91	15.33 ± 4.37	24.13 ± 1.88		
	G/G = 5 (19)	14.63 ± 2.63	17.60 ± 7.67	23.60 ± 1.95		
DBH (rs1611115)	C/C = 17 (65)	14.58 ± 1.59	17.18 ± 5.36	24.82 ± 2.27	0.17	1
	C/T = 9 (35)	16.30 ± 2.23	15.22 ± 5.67	22.78 ± 1.92		
	T/T = 0 (0)	N/A ± N/A	N/A ± N/A	N/A ± N/A		
DRD1 (rs4532)	T/T = 13 (50)	16.03 ± 1.68	15.31 ± 5.06	24.15 ± 2.76	0.35	0.10
	T/C = 8 (31)	14.55 ± 2.40	17.38 ± 6.97	23.63 ± 2.00		
	C/C = 5 (19)	13.93 ± 0.87	18.20 ± 3.70	24.80 ± 1.79		
DRD2 (rs1076560)	C/C = 15 (58)	14.97 ± 2.08	15.00 ± 5.64	23.53 ± 2.45	0.23	1
	C/A = 10 (38)	15.35 ± 1.96	18.30 ± 4.81	25.10 ± 2.02		
	A/A = 1 (4)	16.56 ± N/A	21.00 ± N/A	23.00 ± N/A		
DRD3 (rs6280)	T/T = 11 (42)	14.95 ± 1.93	16.82 ± 5.72	23.73 ± 2.65	0.33	0.67
	T/C = 13 (50)	15.07 ± 1.79	16.77 ± 5.12	24.62 ± 2.22		
	C/C = 2 (8)	17.09 ± 3.71	13.00 ± 8.49	23.00 ± 0.00		

Note. Technical scores presented in mean ± standard deviation. HWE = Hardy-Weinberg equilibrium; MAF = Minor allele frequency.

Table 6.3. Descriptive statistics of subjective technical scores.

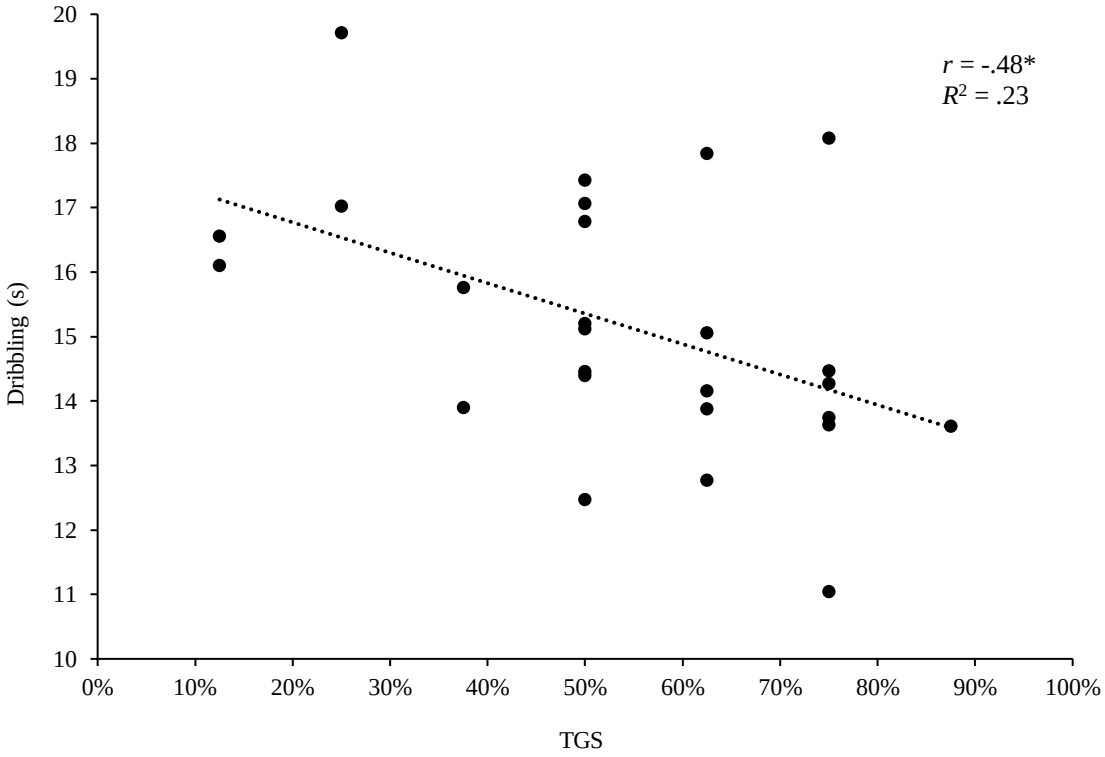
Gene (SNP)	Genotype = *n* (%)	Dribbling	Passing	Shooting	MAF	HWE
Overall		3.06 ± 0.46	2.87 ± 0.45	2.83 ± 0.46	N/A	N/A
ADBR2 (rs1042714)	C/C = 10 (37)	3.24 ± 0.46	2.94 ± 0.43	3.06 ± 0.58	0.44	0.24
	C/G = 10 (37)	2.95 ± 0.47	2.74 ± 0.55	2.57 ± 0.25		
	G/G = 7 (26)	2.95 ± 0.44	2.94 ± 0.34	2.90 ± 0.33		
ACE (rs4341)	C/C = 8 (30)	3.21 ± 0.49	2.92 ± 0.50	2.80 ± 0.38	0.46	1
	C/G = 13 (48)	3.05 ± 0.48	2.80 ± 0.46	2.97 ± 0.60		
	G/G = 6 (22)	2.84 ± 0.35	2.90 ± 0.42	2.64 ± 0.15		
BDNF (rs6265)	C/C = 19 (70)	2.99 ± 0.44	2.93 ± 0.48	2.84 ± 0.44	0.15	1
	C/T = 8 (30)	3.22 ± 0.50	2.71 ± 0.38	2.83 ± 0.53		
	T/T = 0 (0)	N/A ± N/A	N/A ± N/A	N/A ± N/A		
COMT (rs4680)	G/G = 8 (30)	2.92 ± 0.46	2.83 ± 0.42	2.95 ± 0.53	0.46	1
	G/A = 13 (48)	3.12 ± 0.51	2.86 ± 0.49	2.81 ± 0.49		
	A/A = 6 (22)	3.12 ± 0.38	2.93 ± 0.49	2.74 ± 0.30		
DBH (rs1611115)	C/C = 18 (67)	3.08 ± 0.49	2.83 ± 0.42	2.85 ± 0.50	0.19	1
	C/T = 8 (30)	3.04 ± 0.46	2.91 ± 0.57	2.79 ± 0.41		
	T/T = 1 (4)	2.83 ± N/A	3.17 ± N/A	2.88 ± N/A		
DRD1 (rs4532)	T/T = 8 (30)	3.01 ± 0.46	2.97 ± 0.42	2.99 ± 0.57	0.39	0.22
	T/C = 17 (63)	3.05 ± 0.48	2.85 ± 0.49	2.79 ± 0.42		
	C/C = 2 (7)	3.33 ± 0.47	2.56 ± 0.15	2.63 ± 0.18		
DRD2 (rs1076560)	C/C = 16 (59)	3.14 ± 0.47	2.83 ± 0.45	2.78 ± 0.38	0.22	1
	C/A = 10 (37)	2.99 ± 0.44	3.00 ± 0.42	2.97 ± 0.56		
	A/A = 1 (4)	2.44 ± N/A	2.11 ± N/A	2.33 ± N/A		
DRD3 (rs6280)	T/T = 11 (41)	3.00 ± 0.33	2.93 ± 0.48	2.70 ± 0.30	0.37	1
	T/C = 12 (44)	2.95 ± 0.47	2.85 ± 0.49	2.82 ± 0.47		
	C/C = 4 (15)	3.57 ± 0.51	2.74 ± 0.30	3.25 ± 0.65		

Note. Technical scores presented in mean ± standard deviation. HWE = Hardy-Weinberg equilibrium; MAF = Minor allele frequency.

Table 6.4. Objective simple regression analysis.

Gene (SNP)	Model	Skill	*B*	*SE B*	β	*t*	*p*
ADBR2 (rs1042714)	G/G vs. G/C-C/C	Dribbling	1.46	0.58	0.74	2.51	**.019***
		Passing	0.46	2.14	0.08	0.21	.833
		Shooting	-1.92	0.95	-0.82	-2.03	.054
ACE (rs4341)	G/G vs. G/C-C/C	Dribbling	-0.62	0.67	-0.32	-0.93	.360
		Passing	0.06	2.22	0.01	0.03	.978
		Shooting	-0.49	1.06	-0.21	-0.47	.645
BDNF (rs6265)	C/C vs. C/T-T/T	Dribbling	0.21	0.63	0.11	0.33	.741
		Passing	0.02	2.07	0.00	0.01	.994
		Shooting	2.27	0.87	0.97	2.60	**.016***
COMT (rs4680)	G/G vs. G/A-A/A	Dribbling	0.42	0.80	0.21	0.52	.606
		Passing	-0.76	2.63	-0.14	-0.29	.775
		Shooting	-1.07	1.24	-0.46	-0.86	.396
DBH (rs1611115)	C/C vs. C/T-T/T	Dribbling	-1.27	0.59	0.64	2.15	**.042***
		Passing	1.01	2.09	-0.19	-0.48	.633
		Shooting	1.96	0.92	-0.84	-2.13	**.044***
DRD1 (rs4532)	T/T vs. T/C-C/C	Dribbling	1.37	0.54	0.69	2.55	**.018***
		Passing	-1.70	1.95	-0.31	-0.87	.395
		Shooting	0.20	0.95	0.08	0.21	.839
DRD2 (rs1076560)	C/C vs. C/A-A/A	Dribbling	-0.82	0.59	0.42	1.39	.177
		Passing	-2.97	1.91	0.55	1.56	.133
		Shooting	-1.30	0.92	0.55	1.41	.173
DRD3 (rs6280)	T/T vs. T/C-C/C	Dribbling	-0.37	0.61	-0.19	-0.62	.544
		Passing	0.52	1.99	0.10	0.26	.795
		Shooting	-0.68	0.94	-0.29	-0.72	.480

Note. Bold values and * highlight statistical significance at $p < .05$. *B* = unstandardised beta; *SE B* = standard error; β = standardised beta.



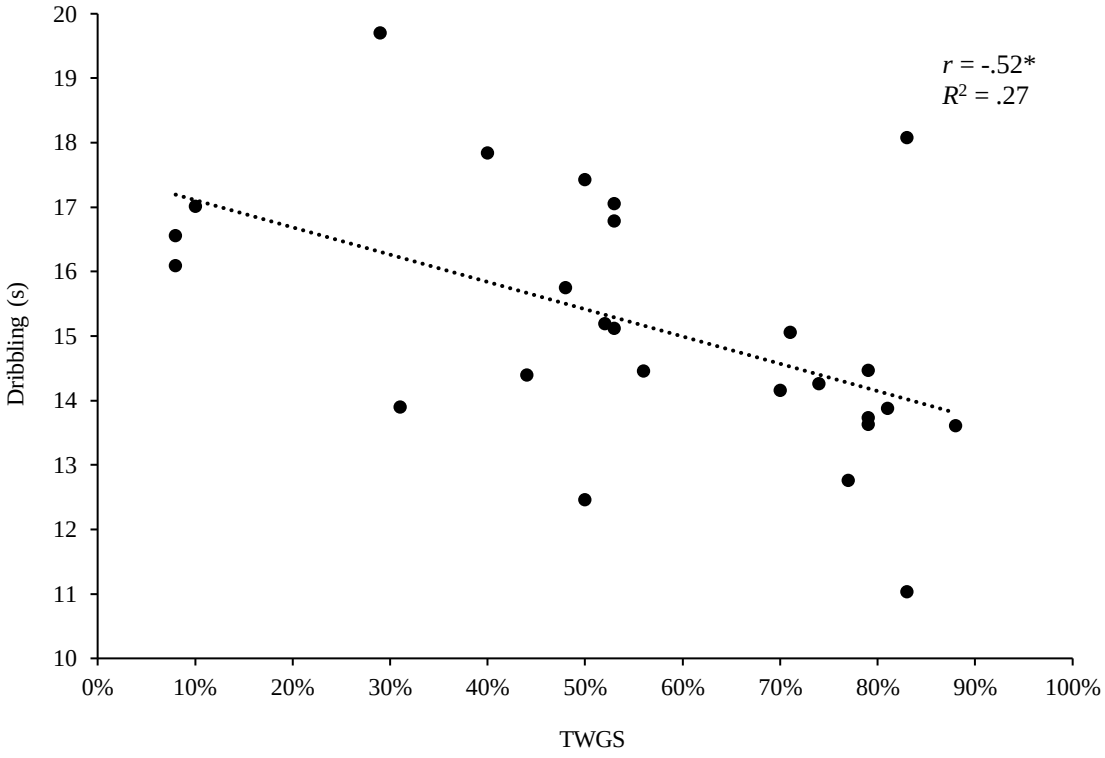
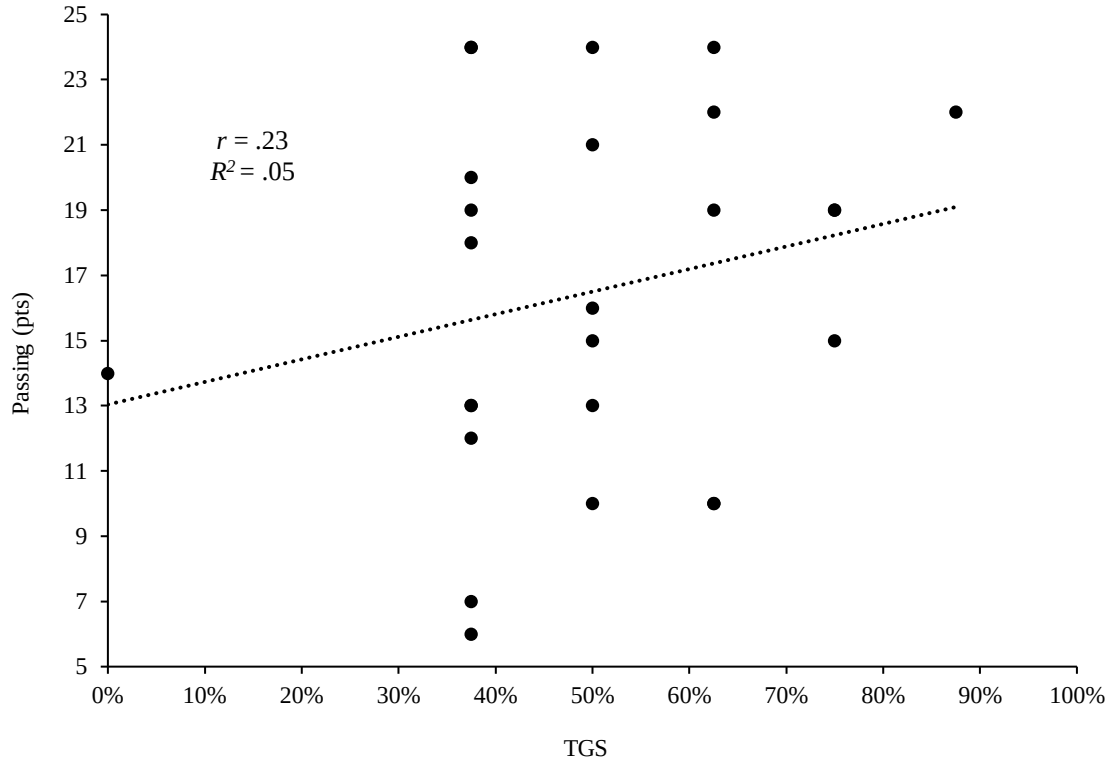


Figure 6.1. Objective total genotype score (TGS) and total weighted genotype score (TWGS) correlations with dribbling. * Statistically significant at $p < .05$.



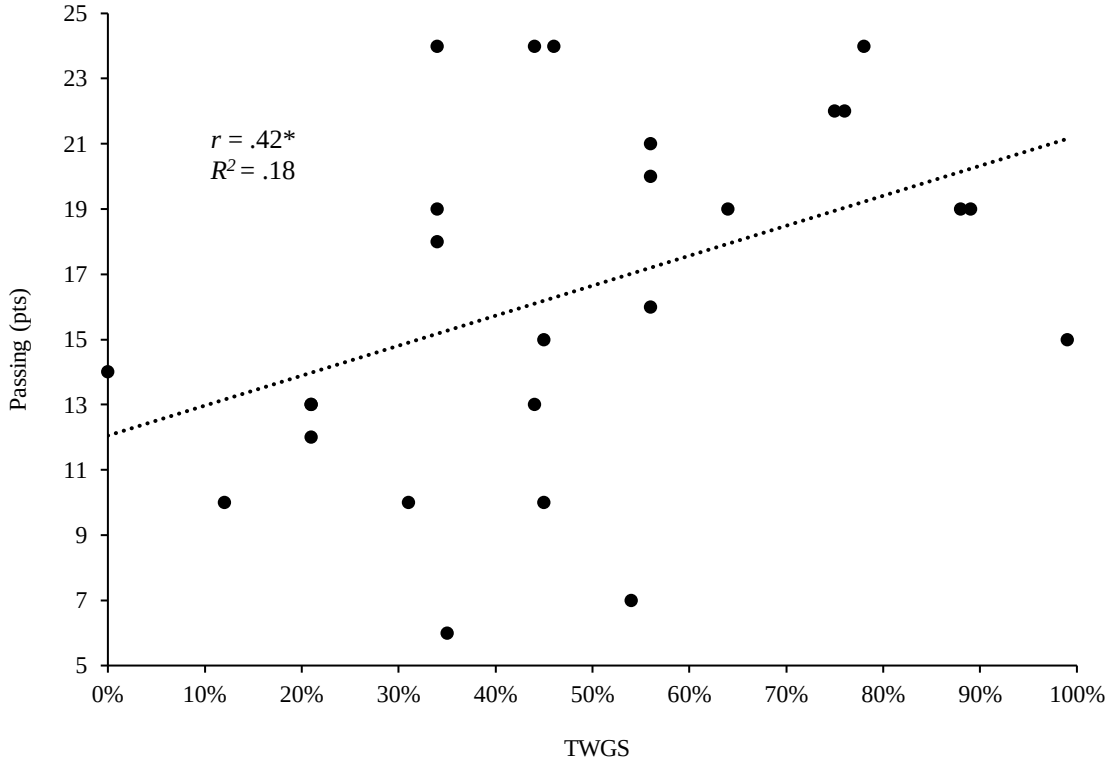
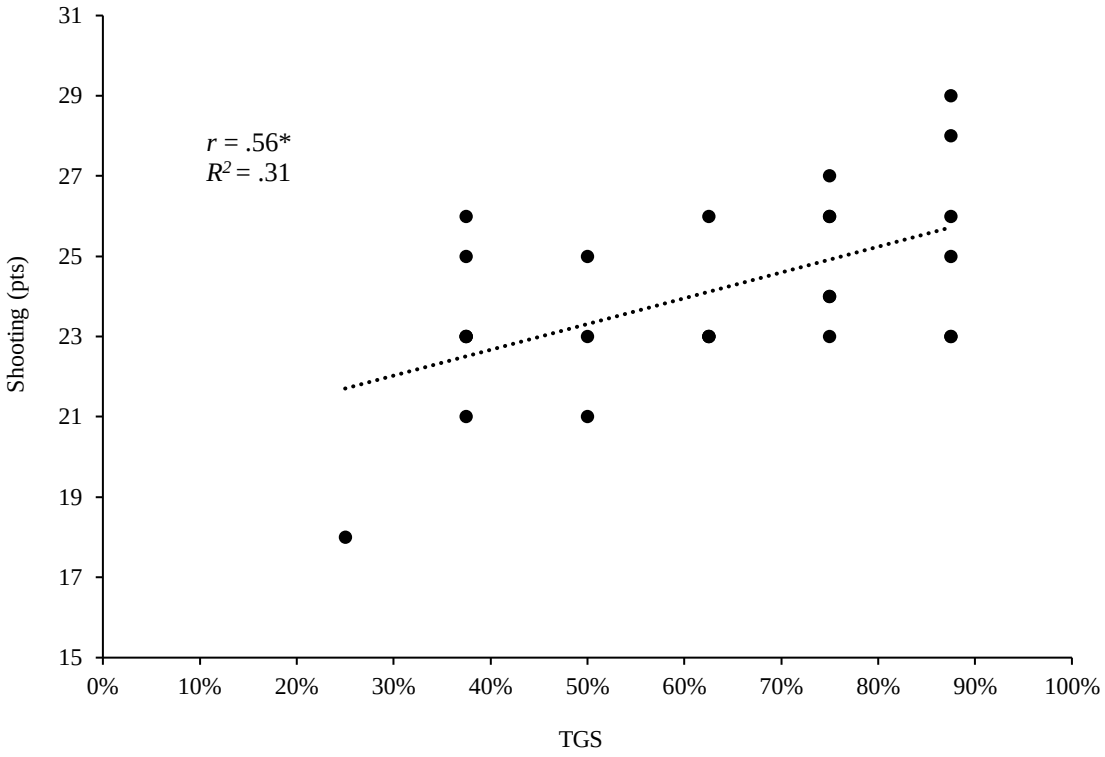


Figure 6.2. Objective total genotype score (TGS) and total weighted genotype score (TWGS) correlations with passing. * Statistically significant at $p < .05$.



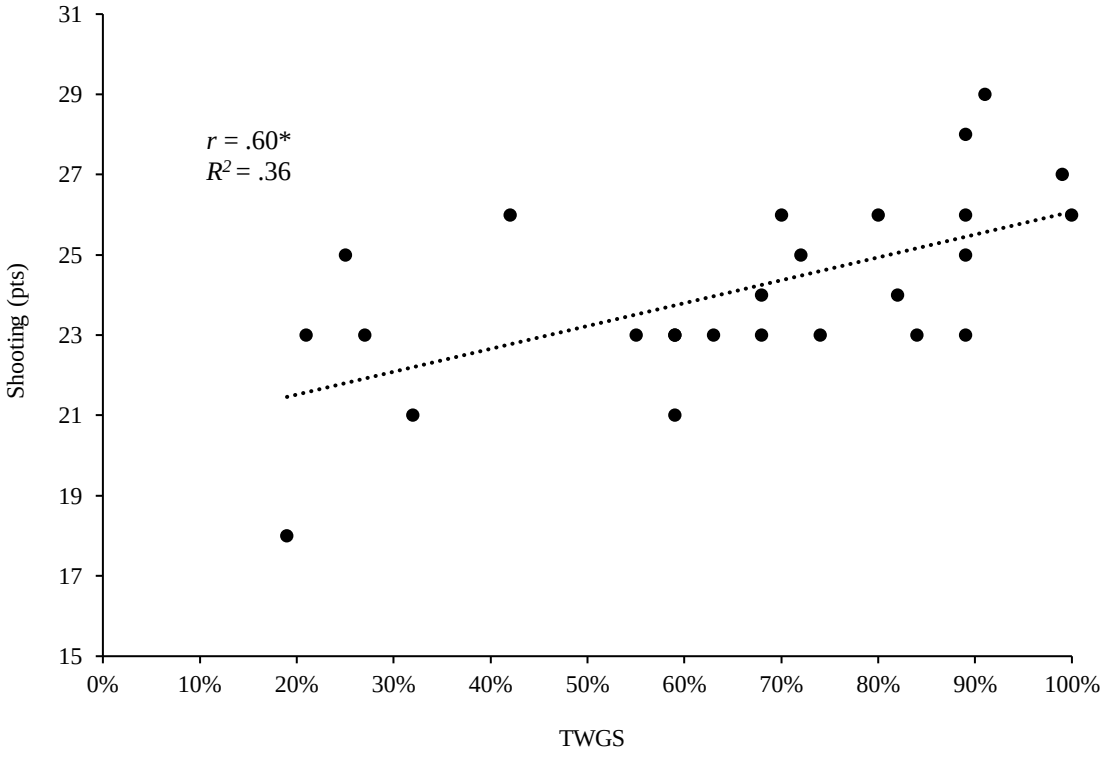
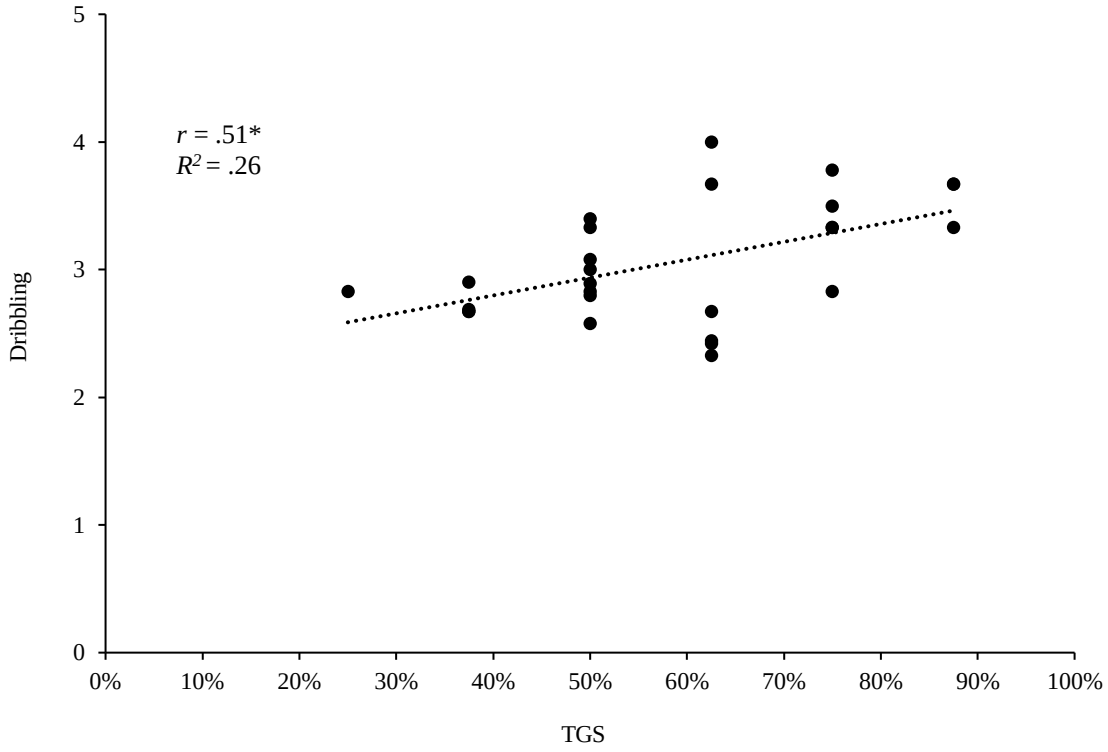


Figure 6.3. Objective total genotype score (TGS) and total weighted genotype score (TWGS) correlations with shooting. * Statistically significant at $p < .05$.

Table 6.5. Subjective simple regression analysis.

Gene (SNP)	Model	Skill	*B*	*SE B*	β	*t*	*p*
ADBR2 (rs1042714)	C/C vs. C/G-G/G	Dribbling	0.29	0.18	0.63	1.62	.118
		Passing	0.12	0.18	0.27	0.66	.514
		Shooting	0.35	0.17	0.77	2.05	.051
ACE (rs4341)	G/G vs. G/C-C/C	Dribbling	-0.28	0.21	-0.61	-1.34	.194
		Passing	0.04	0.21	0.08	0.18	.859
		Shooting	-0.25	0.21	-0.54	-1.17	.254
BDNF (rs6265)	C/C vs. C/T-T/T	Dribbling	-0.22	0.19	-0.48	-1.15	.261
		Passing	0.22	0.19	0.49	1.18	.249
		Shooting	0.01	0.20	0.03	0.07	.946
COMT (rs4680)	G/G vs. G/A-A/A	Dribbling	-0.19	0.19	-0.42	-0.99	.332
		Passing	-0.05	0.19	-0.12	-0.28	.784
		Shooting	0.16	0.19	0.36	0.85	.405
DBH (rs1611115)	C/C vs. C/T-T/T	Dribbling	0.07	0.19	0.15	0.36	.721
		Passing	-0.10	0.19	-0.22	-0.54	.593
		Shooting	0.05	0.19	0.11	0.25	.802
DRD1 (rs4532)	T/T vs. T/C-C/C	Dribbling	-0.07	0.20	-0.15	-0.35	.732
		Passing	0.15	0.19	0.33	0.77	.448
		Shooting	0.22	0.19	0.48	1.13	.268
DRD2 (rs1076560)	C/C vs. C/A-A/A	Dribbling	0.21	0.18	0.44	1.14	.265
		Passing	-0.10	0.18	-0.21	-0.53	.599
		Shooting	-0.13	0.18	-0.29	-0.72	.477
DRD3 (rs6280)	T/T vs. T/C-C/C	Dribbling	-0.11	0.18	-0.24	-0.60	.553
		Passing	0.11	0.18	0.25	0.63	.536
		Shooting	-0.23	0.18	-0.49	-1.27	.215

Note. *B* = unstandardised beta; *SE B* = standard error; β = standardised beta.



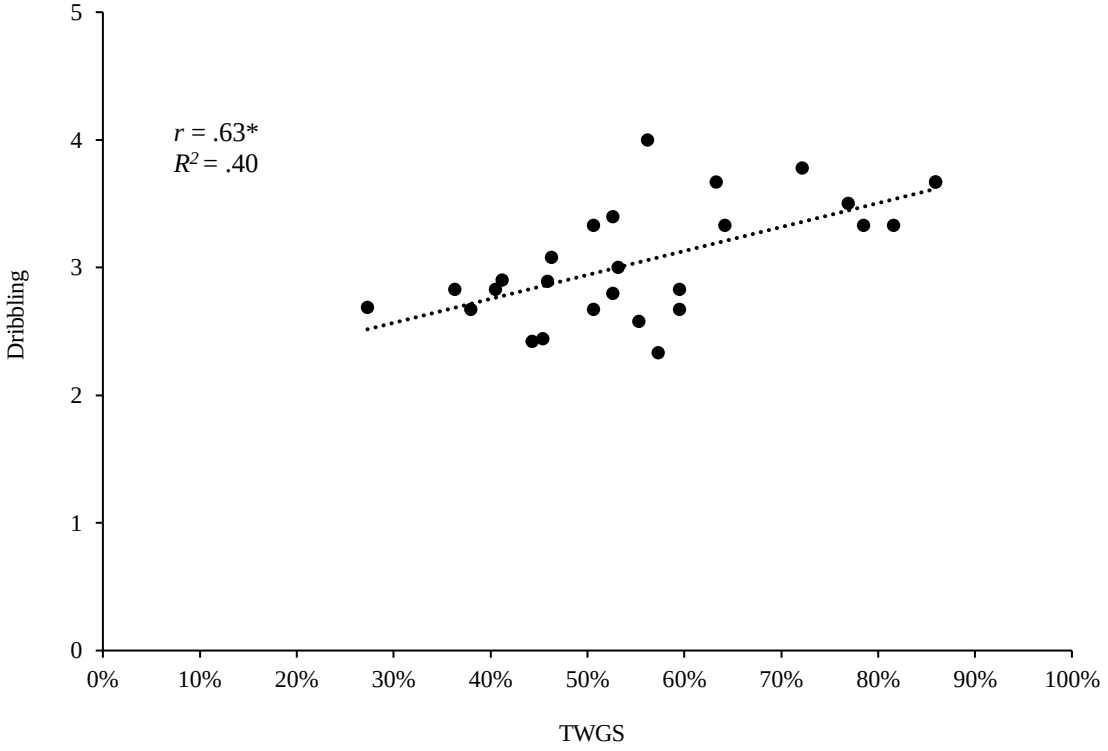
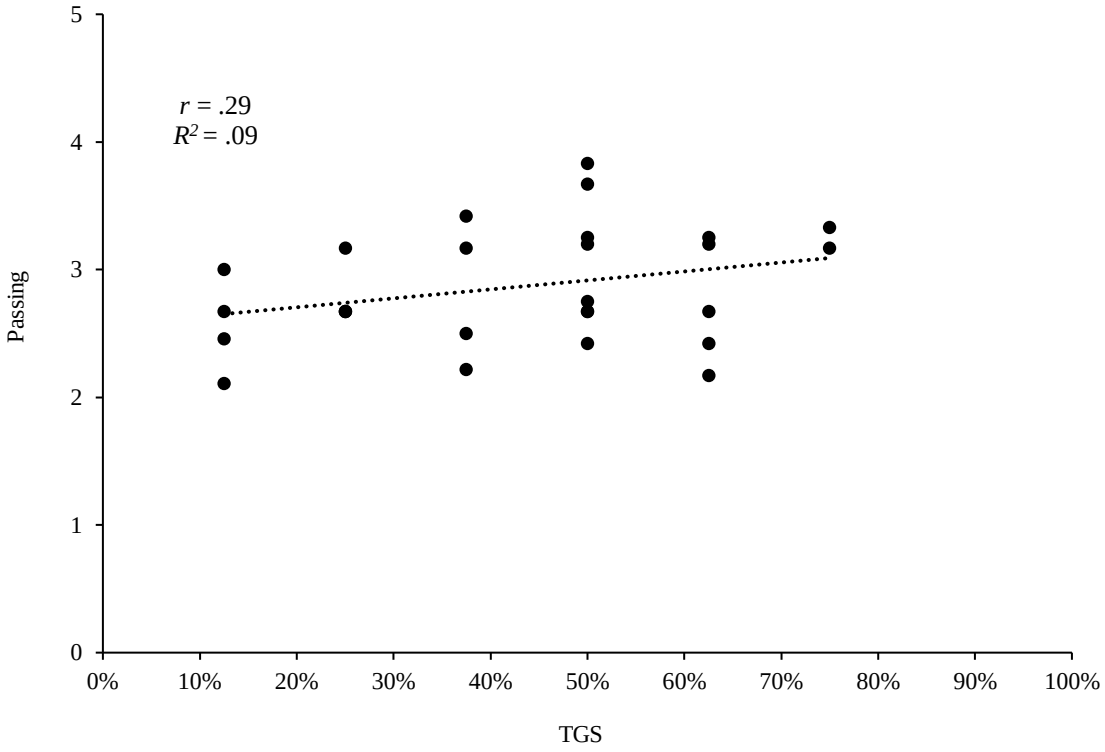


Figure 6.4. Subjective total genotype score (TGS) and total weighted genotype score (TWGS) correlations with dribbling. * Statistically significant at $p < .05$.



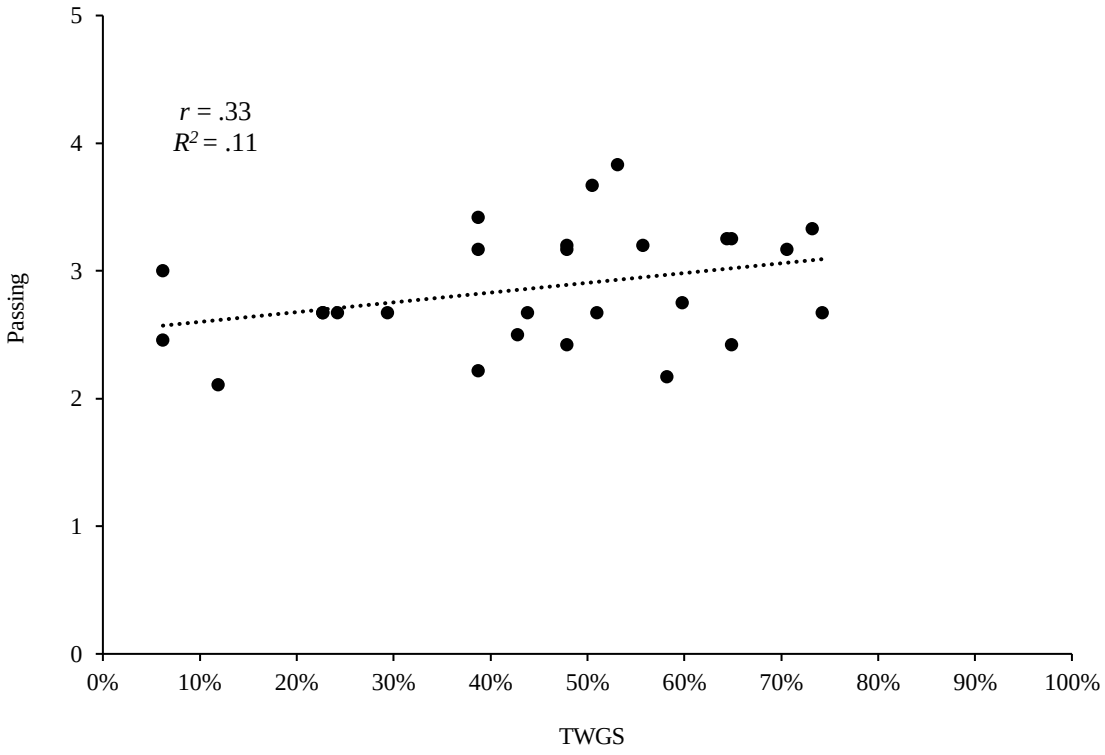
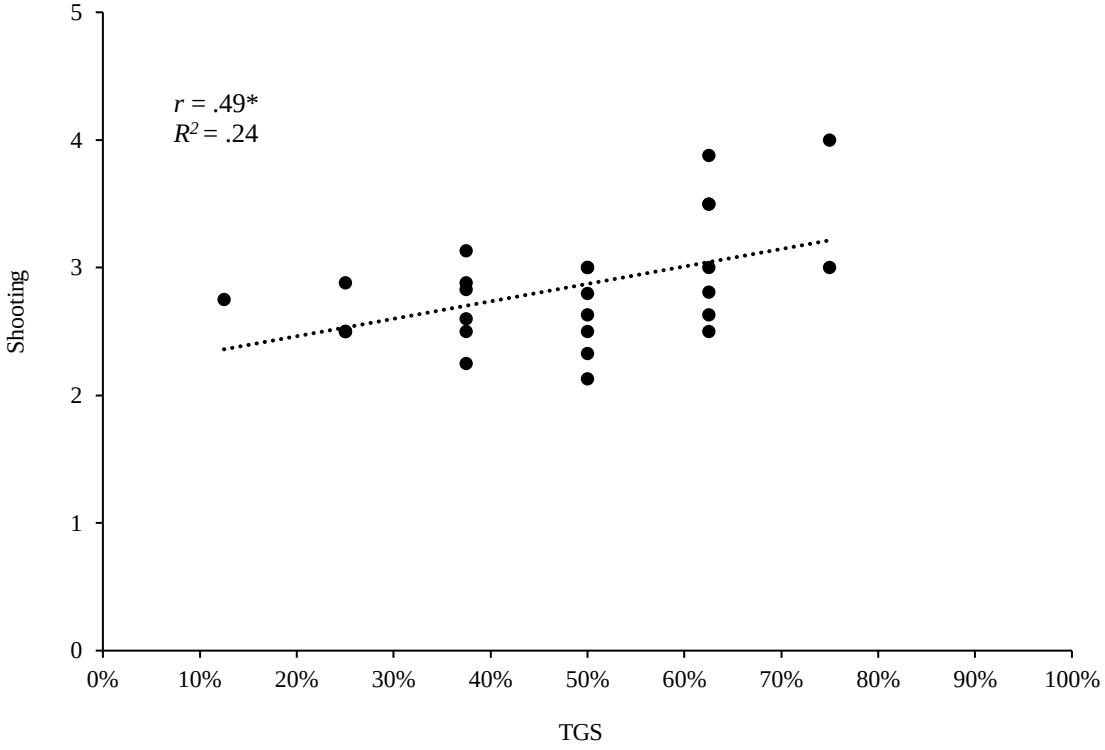


Figure 6.5. Subjective total genotype score (TGS) and total weighted genotype score (TWGS) correlations with passing. * Statistically significant at $p < .05$.



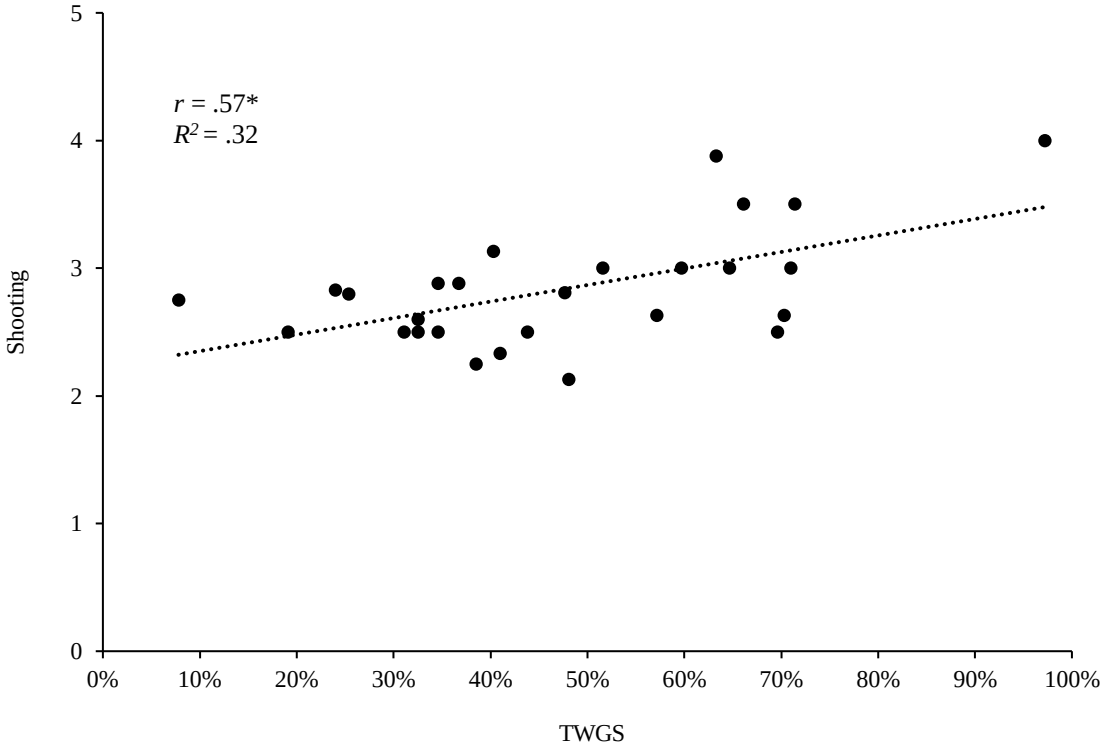


Figure 6.6. Subjective total genotype score (TGS) and total weighted genotype score (TWGS) correlations with shooting. * Statistically significant at $p < .05$.

6.4 Discussion

This preliminary study investigated the association of eight SNPs, both individually and collectively, with objective and subjective assessments of technical capabilities in youth football players. In doing so, this study has shown for the first time: (a) genetic variations may be associated with the technical capabilities of youth football players, (b) the genetic variants that may be responsible for inter-individual variability in technical skill, (c) phenotypic distinctions in the characteristics of objective and subjective assessments may alter allelic associations, and (d) advantageous genotypes may have an additive effect on technical performance in football, irrespective of assessment method. As such, these preliminary findings may have important implications for future football genomic studies.

The single SNP analysis revealed four SNPs associated with objective dribbling and shooting assessments, but there were no associations with the objective passing assessments. This may reflect the different phenotypic characteristics of the tests and/or the much wider distribution of scores in the passing test compared to dribbling and shooting, which would have decreased statistical power in the passing test. The SNPs associated with either of the dribbling and shooting objective assessments were: *ADBR2* (rs1042714), *BDNF* (rs6265), *DBH* (rs1611115), and *DRD1* (rs4532). Studies in this area, with this type of unique sample, are typically underpowered, which can make it difficult to clearly make conclusions about what these findings mean. However, in early stages of development in a field, informed speculation based on prior knowledge may be important for informing future work. As a result, we have made informed speculation about our findings as a way of guiding subsequent work in this area.

The C allele of *ADBR2* (rs1042714) was associated with a faster completion time in the slalom dribble test and exhibited the largest effect of any SNP on this phenotype. These results correspond with those of Jacob et al. (2018), who reported that C allele carriers performed

significantly better during a skill assessment in AFL. The findings of both studies indicate that *ADBR2* (rs1042714) may play a role in motor control and/or development. To the authors' knowledge, the association of *ADBR2* (rs1042714) with motor skills has yet to be investigated outside of the sport domain. The G allele has, however, been associated with an increased risk of autism (Cheslack-Postova et al., 2007). Although speculative, the proposed underpinning mechanisms of this association may help elucidate the results of this study, as autism is in part characterised by deficits in skill acquisition (Moraes et al., 2017). The *ADRB2* gene encodes the beta-2-adrenergic receptor, which is widely expressed in the brain as part of the catecholamine system and acts as a receptor for adrenaline, noradrenaline, and dopamine (Brodde, 2008; Dohlman et al., 1991). The *ADRB2* (rs1042714) SNP modulates receptor activity and sensitivity (Cagliani et al., 2009; Johnson 2006), whereby the G allele is associated with increased responsiveness to ligand and delayed desensitisation and downregulation. As the G allele is more responsive, it may increase vulnerability to the associated effects of overstimulation (i.e., altered cell replication, differentiation, morphology and distribution), causing neurodevelopmental disorders by modifying neural architecture (Cheslack-Postova et al., 2007), and consequently deficits in technical ability in football. Other research has also shown that *ADBR2* (rs1042714) may be associated with changes in white matter and cognitive functions, although specific allelic associations have been inconsistent (Lyall et al., 2013). As such, further research with larger samples is required to validate the association of *ADBR2* (rs1042714) with technical ability in football as current mechanistic explanations are speculative.

The C/C genotype of *BDNF* (rs6265) was associated with a higher score in the shooting accuracy test and exhibited the largest effect of any SNP on this phenotype. These results are in accordance with a plethora of research that suggests the T allele is associated with a decreased motor learning capacity. The *BDNF* gene encodes for the BDNF protein, which

influences cortical synaptic plasticity (Akaneya et al., 1997). Carriers of the *BDNF* (rs6265) T allele have shown lower increases in the amplitude of motor-evoked potentials and motor map reorganisation following motor training (Kleim et al., 2006). Moreover, following transcranial magnetic stimulation, there was no change in the neurological excitability of individuals possessing the T allele, whereas there was a 67% increase in C allele carriers (Missitzi et al., 2011). This suggests T allele carriers may have a lower motor learning adaptation capacity due to less neurobiological excitability, which may be related to altered cortical synaptic plasticity. Indeed, it has been reported that the T allele produces a lower activity-dependent release and recruitment of BDNF in neurons, altered glutamatergic and GABAergic synaptic transmission, and changes of cortical and hippocampal morphology resulting in deficits in learning and memory (Deveci et al., 2020). As such, the association of *BDNF* (rs6265) with technical ability in football may be explained by its potential influence on motor learning.

The C allele of *DRD1* (rs4532) was associated with a faster completion time of the slalom dribble test. The *DRD1* gene encodes the DRD1 protein, which helps regulate synaptic dopamine levels by subserving dopamine neurotransmission across motor cortices and the basal ganglia (Gu, 2002; Hauber, 1998). The *DRD1* (rs4532) SNP is in a 5′ untranslated regulatory region that may affect mRNA stability and translation, as the C allele has been associated with several disorders associated with increased brain dopamine neurotransmission (Cichon et al., 1994; Comings et al., 1997). Increased dopamine levels have been found to promote motor learning and motor cortex plasticity, so long as an excess of dopamine is not accumulated (Matsumoto et al., 1997; Meintzschel & Ziemann, 2006). Indeed, when *DRD1* (rs4532) was incorporated into a TGS, with the C allele classified as an advantageous genotype, the TGS was associated with greater motor learning (Pearson-Fuhrhop et al., 2013). As such, the findings of this study suggest a role for *DRD1* (rs4532) on technical ability in football due to the potential association with dopamine levels and consequently motor learning.

The C/C genotype of *DBH* (rs1611115) was associated with a faster completion time of the slalom dribble test and a higher score in the shooting accuracy test. The *DBH* gene encodes the DBH enzyme which catalyses the oxidative hydroxylation of dopamine to norepinephrine (Gonzalez-Lopez & Vrana, 2020). DBH enzymatic activity modulates norepinephrine levels and influences executive and motor function (Mustapic et al., 2013). It has been estimated that the *DBH* (rs1611115) SNP explains between 30 to 50% of the variance in DBH activity, with the T allele associated with decreased DBH activity (Zabetian et al., 2001). Since DBH plays an important role in the metabolism of dopamine and norepinephrine, decreased DBH activity may facilitate structural or functional neuronal damage, and consequently neurodevelopmental and neurodegenerative disorders (Gonzalez-Lopez & Vrana, 2020). Therefore, as the C allele may be associated with higher DBH activity and consequently improved motor function, this mechanism may explain the association of *DBH* (rs1611115) with technical ability in football.

There were no associations between a single SNP and the subjective coach ratings of any technical skill. This may in part be due to the specific characteristics of a different cohort, but it is also likely because of the unique criteria coaches use to judge player and technical ability. Unfortunately, coaches have found it difficult not only to describe technical proficiency, but also explicitly report the criteria they use to assess athlete abilities (Roberts et al., 2019). Indeed, as a recent review highlighted (see Lath et al., 2021), there is a lack of knowledge and clear understanding of how coaches subjectively assess the abilities of players throughout the literature, and as a result, the underlying mechanism(s) remains unknown. Therefore, the specific phenotypic characteristics coaches evaluate in players to infer technical ability, may be subtly different from those that facilitate improved performance in objective tests.

Research has indicated that objective and subjective assessments of player ability in football have similar prognostic validity (Honer et al., 2021; Sieghartsleitner et al., 2019). However, the levels of agreement between subjective and objective assessments of current performance have only been confirmed using physical assessments (Dugdale et al., 2020). Moreover, Dugdale et al. (2020) reported that coaches' subjective assessments of physical assessments only corresponded with the extremities (i.e., highest and lowest objective performances), suggesting that subjective assessments may lack sensitivity when discriminating between players of similar ability. Although these findings are yet to be replicated with technical assessments, they may help explain the non-significant results of this study due to the narrow distribution of subjective scores.

The TGSs and TWGSs were associated with several objective and subjective technical assessments. This suggests the technical skills of dribbling, passing, and shooting in football are polygenic traits and the SNPs in this study have an additive effect on each skill. This finding is in accordance with a previous investigation on the collective influence of genetic polymorphisms on motor learning, which found advantageous alleles of genetic variants associated with dopamine neurotransmission influence the motor system in an additive manner (Pearson-Fuhrhop et al., 2013). This study also showed weighting SNPs by their partial influence improved the relationship between each GS model and the respective technical phenotype under investigation. This finding is congruent with previous research comparing unweighted and weighted TGSs using a similar approach. For instance, Massidda et al. (2014) found a TWGS explained 10% and 15% more of the variance than a TGS in countermovement jump and squat jump height, respectively. In this study, the TWGSs explained 4-13% and 2-14% more of the variance than the TGSs in the objective and subjective assessments, respectively. These findings suggest each advantageous allele of a given SNP may have a different degree of influence on a specific technical phenotype. Moreover, these findings

showcase that dribbling, passing, and shooting all have subtly different phenotypic characteristics, resulting in each having unique advantageous genotypes and polygenic profiles. However, some SNPs also have a pleiotropic effect, whereby they influence separate technical phenotypes in a similar manner (e.g., *DBH* C/C genotype on objective dribbling and shooting performance).

This study has several limitations which should be acknowledged when interpreting its results. The main limitations derive from the size of the samples and the number of association tests performed without adjustment for multiple comparisons. However, the sample sizes used in this study are similar to that of other preliminary studies in this research area (e.g., Jacob et al. 2018) and indeed football genetics as a whole (see McAuley, Hughes, et al., 2021a for a review). Furthermore, adjustments for multiple comparisons in exploratory research are not recommended (see Althouse, 2016), as reducing Type 2 errors is the priority to ensure an important novel association is not missed, which can be confirmed or rejected in subsequent higher-powered studies. Nevertheless, the observed associations could be false positives or might only be specific to the samples under investigation, and consequently may not be generalisable to other football cohorts. As such, further research is required with independent and larger football cohorts to replicate and assess the external validity of these results before practical applications can be recommended. Building this research base with studies using transparent methodologies is important so they can contribute to research synthesis approaches in the future to draw more valid and reliable conclusions (McAuley, Baker, et al., 2022). There are many other factors that may influence performance in technical assessments that could not be controlled for. For instance, recent research has shown that maturation status is associated with distinct genetic profiles in youth football players (Murtagh et al., 2020). Capturing and adjusting for maturation status as well as chronological age and other confounding variables may provide greater context to findings.

6.5 Conclusion

This study indicates inter-individual genetic variation may influence the technical capabilities of youth football players. These findings suggest *ADBR2* (rs1042714), *BDNF* (rs6265), *DBH* (rs1611115), and *DRD1* (rs4532), in isolation, may be significant predictors of dribbling and shooting performance when using objective assessment methods. In addition, the polygenic models of all the SNPs included in this study were shown to be significant predictors of technical capabilities, irrespective of assessment method. As such, these SNPs may prove to be a useful starting point in establishing a genetic profile tool capable of assisting with technical performance assessment and development. However, before the results of this study can be considered for practical applications, replications via higher-powered research designs are necessary.

7 Genetic associations with personality and mental toughness profiles of English academy football players: An exploratory study

McAuley, A. B. T., Hughes, D. C., Tsaprouni, L. G., Varley, I., Suraci, B., Baker, J., Herbert, A. J., & Kelly, A. L. (2022). Genetic associations with personality and mental toughness profiles of English academy football players: An exploratory study. *Psychology of Sport and Exercise*, *61*, 102209. https://doi.org/10.1016/j.psychsport.2022.102209

Abstract

Psychological characteristics influence the performance of youth football players and are significant predictors of development and success at adulthood. Although genetic factors may explain a considerable portion of inter-individual differences in psychological traits, psychogenetic research in football is scarce. As such, the purpose of this study was to examine the association of ten single nucleotide polymorphisms (SNPs) with personality and mental toughness profiles of academy football players. Seventy-three male under-12 to under-18 football players from a Category 3 English academy were genotyped for ten SNPs. Personality and mental toughness were assessed using a 50-item IPIP Big Five personality traits questionnaire and the Mental Toughness Index, respectively. Simple linear regression was used to analyse individual SNP associations with personality dimensions and mental toughness, whereas both unweighted and weighted total genotype scores (TGSs; TWGSs) were computed to measure the combined influence of all SNPs. There was a significant association between *DRD3* (rs167771) and agreeableness, where A/A homozygotes scored higher than G allele carriers. TGSs and/or TWGSs were significantly correlated with mental toughness and each

personality dimension except openness, explaining between 3-17% of the variance. The results of this study suggest psychological characteristics of youth football players are partly determined by genetic factors.

7.1 Introduction

Psychological characteristics are now an integral component of multidimensional athlete development models in football (soccer) (Vaeyens et al., 2008; Williams et al., 2020). Early research in football indicated higher performing youth football players possessed superior psychological capacities that facilitated greater development than their lower performing counterparts (Williams & Reilly, 2000). Psychological attributes have been researched frequently over the last two decades and have not only been regularly associated with current performance levels in youth football players, but are also significant predictors of development and success in adulthood (Murr et al., 2018). As a result, psychological aspects have been integrated into many countries' multidisciplinary long-term youth development strategies (e.g., the Elite Player Performance Plan in England; Premier League, 2011). However, at present, research on the psychological aspects of performance in football is scarce compared to other multidimensional components, such as physiological and technical (Sarmento et al., 2018).

Psychological research in football has identified a wide range of important psychological characteristics, which have been associated with objective performance metrics as well as through subjective coach, player, and scout perceptions (Williams et al., 2020). The predominant method of assessing psychological characteristics has been through self-report validated questionnaires/scales (Musculus & Lobinger, 2018). Several recent systematic reviews on football-specific research have synthesised the findings of these studies from both psychological (e.g., Gledhill et al., 2017; Murr et al., 2018; Ivarsson et al., 2020) and multidisciplinary (e.g., Bergkamp et al., 2019; Sarmento et al., 2018; Williams et al., 2020) perspectives. Overall, and similar to other multidimensional predictors, studies have generally reported small to moderate effect sizes for psychological variables on current and future

performance in football. Some of the more strongly supported psychological variables include: (a) achievement goal orientation, (b) achievement motives, (c) commitment, (d) concentration, (e) coping strategies, (f) effort, (g) motivation, (h) resilience, (i) self-determination, (j) self-regulation, and (k) task orientation (Murr et al., 2018; Ivarsson et al., 2020).

In psychological research, there are umbrella terms which encompass many of the psychological variables outlined, such as personality and mental toughness (Lin et al., 2017). Personality is most commonly assessed in contemporary sport research using the Big Five/Five Factor model, which contends that there are five main personality dimensions (i.e., extraversion, agreeableness, openness, conscientiousness, and neuroticism) that embody several more specific psychological facets (McCrae & John, 1992). More specifically, extraversion assesses the quantity and intensity of interpersonal interactions, agreeableness assesses individuals' concern for cooperation and social harmony, openness assesses individuals' tendency to seek out new experiences, conscientiousness assesses organisation and goal-directed behaviour, and neuroticism assesses the degree to which individuals are prone to emotional instability (Allen et al., 2013). Mental toughness on the other hand has been defined as a collection of psychological resources that facilitate an individual's capacity to produce reliable objective and subjective performance in the presence of varying situational demands (Gucciardi et al., 2015). Unsurprisingly, several studies have reported significant associations between personality traits, mental toughness, and performance in sport (Allen et al., 2013; Liew et al., 2019). In general, more successful athletes tend to score lower on neuroticism and higher on mental toughness (Benítez-Sillero et al., 2021; Piepiora, 2021; Steca et al., 2018).

To assess the underpinning mechanisms of psychological variables, behavioural scientists have investigated the extent to which genetic and environmental factors influence

inter-individual differences (Chabris et al., 2015). These studies identified a sizable genetic component in both personality and mental toughness (Allen et al., 2013; Lin et al., 2017; Power & Pluess, 2015). Specifically, the heritability estimate for overall personality and mental toughness is approximately 50%, with the sub-components of each generally ranging from 35% to 65% (Horsburgh et al., 2009). There is considerable evidence supporting the influence of genetics on psychological variables. For example, Turkheimer (2000) noted in a seminal article that the "Three Laws of Behaviour Genetics" are: (a) all human behavioural traits are heritable, (b) the effect of being raised in the same family is smaller than the effect of genes, and (c) a substantial portion of the variation in complex human behavioural traits is not accounted for by the effects of genes or families. These suggestions are further supported by a meta-analysis encompassing 50 years of twin research reporting an average heritability estimate across all behavioural and physical traits of 49% (Polderman et al., 2015).

Grounded on the amalgamation of a vast amount of empirical evidence from molecular genetic research, a "Fourth Law of Behaviour Genetics" was proposed by Chabris et al. (2015). Namely, a typical human behavioural trait is associated with very many genetic variants, each of which accounts for a very small percentage of the behavioural variability. As such, analogous to physical phenotypes, the genetic architecture of psychological phenotypes is highly polygenic (i.e., many different genetic variants influence each trait) and pleiotropic (i.e., each genetic variant influences many different traits). Although, in comparison to physical phenotypes, psychological phenotypes may be *more* polygenic based on the number and mean effect sizes of the associated genetic variants identified in contemporary research (Chabris et al., 2015). Most of these common genetic variants (i.e., polymorphisms) are within genes in the serotonergic and dopaminergic systems (Ausmees et al., 2021; Balestri et al., 2014).

However, whilst there is considerable research on these and other variants, the validity of their associations with psychological traits requires further research (Karlsson Linnér et al., 2019; Sanchez-Roige et al., 2018; Strawbridge et al., 2018).

Within a football-specific context, there is very limited psychogenetic research. For instance, a recent systematic review of genetic association research in football only identified three studies involving football players (see McAuley, Hughes, et al., 2021a). These studies reported significant genetic associations with overall athlete status (Filonzi et al., 2015) and specific psychological phenotypes such as anxiety, depression, impulse control, and neuroticism (Cochrane et al., 2018; Petito et al., 2016). However, the cohorts under investigation in these studies consisted of a combination of footballers and athletes from other sports, which limits the implications of their findings to a football-specific context. To the authors' knowledge, there is yet to be a psychogenetic study on footballers in isolation. Accordingly, the aim of this exploratory study was to examine the associations amongst ten psychogenetic single nucleotide polymorphisms (SNPs) with personality and mental toughness profiles of academy football players. Identifying a panel of SNPs associated with relevant psychological phenotypes may enhance athlete development processes in football by enabling more individualised psychological intervention programmes, which may help manage athlete welfare and improve long-term performance.

7.2 Methods

7.2.1 Participants

In total, 73 male under-12 to under-18 (aged 14.31 ± 2.16 years) football players from a Category 3 English academy participated in this study. Prior to the study commencing, informed assent from all players, consent from parents/guardians, and gatekeeper consent from

each academy was collected. Coaches of each age group were present whilst players completed the questionnaires. All experimental procedures were conducted in accordance with the guidelines in the Declaration of Helsinki and ethical approval was granted by Birmingham City University via the Health, Education, and Life Sciences Academic Ethics Committee. This study was conducted in accordance with the recommendations for reporting the results of genetic association studies defined by the STrengthening the REporting of Genetic Association studies (STREGA) Statement (Little et al., 2009).

7.2.2 Personality

Personality was assessed using a 50-item Big Five personality traits questionnaire from the International Personality Item Pool (IPIP) (Goldberg et al., 2006). Each personality trait was measured by the sum of each participant's answers to ten statements using a five-point Likert scale ranging from one (disagree) to five (agree). For example, 'I am the life of the party' (extraversion), 'I am interested in people' (agreeableness), 'I am full of ideas' (openness), 'I am always prepared' (conscientiousness), and 'I get stressed out easily' (neuroticism). Higher scores represent higher levels of each personality trait. Previous studies that applied this tool have reported good internal consistency (Cronbach's alpha) across all trait measurements (i.e., extraversion = 0.87, agreeableness = 0.81, openness = 0.78, conscientiousness = 0.77, and neuroticism = 0.88) (Power & Pluess, 2015).

7.2.3 Mental toughness

Mental toughness was assessed using the Mental Toughness Index (MTI) (Gucciardi et al., 2015). The MTI is an eight-item unidimensional measure which instructs participants to indicate how they typically think, feel, and behave as an athlete using a seven-point Likert scale ranging from one (False, 100% of the time) to seven (True, 100% of the time) (Cooper et al.,

2022; Jones & Parker, 2019). The sum of the eight items yields a mental toughness score that is then used for analysis, with higher scores representing higher levels of mental toughness. Previous studies examining the internal consistency and reliability of the MTI have demonstrated both high Cronbach's alpha (0.90) and composite reliability (0.90) levels (Jones & Parker, 2018).

7.2.4 Genetic procedures

7.2.4.1 Genotyping

Saliva was collected from players via sterile, self-administered buccal swabs, following a minimum of 30 minutes since food or drink ingestion. Saliva samples were safely stored at room temperature and within 36 hours were sent to AKESOgen, Inc. (Peachtree Corners, GA, USA) for DNA extraction. Using Qiagen chemistry, DNA was extracted on an automated Kingfisher FLEX instrument (Thermo Fisher Scientific, Waltham, MA, US). The manufacturers recommended guidelines and procedures were followed throughout. To measure the extracted DNA's quality and quantity, PicoGreen and Nanodrop measurements were taken. Input to the custom testing array occurs at 200 ng in 20 μL. Biomek FXP was used to perform amplification, fragmentation, and resuspension. Hybridisation was performed in a Binder oven at 48 degrees for 24 hours and GeneTitan instrumentation (Thermo Fisher Scientific, Waltham, MA, US) was used to stain and scan the arrays, following the Affymetrix Axiom high throughput 2.0 protocol. Data analysis was then performed using a raw CEL file data input into the Affymetrix Axiom Analysis Suite (Affymetrix, Santa Clara, CA, US). Procedures are in accordance with Pickering et al. (2019).

7.2.4.2 Polymorphism selection

To identify potentially associated polymorphisms, empirical research, review articles, book chapters, and the GWAS catalog (https://www.ebi.ac.uk/gwas/) were examined. Priority was given to GWAS results that were replicated in independent cohorts, followed by candidate gene studies with large homogenous sample sizes which produced similar associations in more than one study. The polymorphisms that were finally included were dependent on the coverage of the microarray and quality control procedures. After an extensive search of the literature, the following ten polymorphisms were selected based on their proposed biological function and relevant associations in previous studies: 5-hydroxytryptamine receptor 2A (*HTR2A*; rs6311), Brain derived neurotrophic factor (*BDNF*; rs6265), Cholinergic receptor muscarinic 2 (*CHRM2*; rs1824024), Catechol-O-methyltransferase (*COMT*; rs4680), Catenin alpha 2 (*CTNNA2*; rs7600563), Dopamine receptor D2 (*DRD2*; rs1800497), Dopamine receptor D3 (*DRD3*; rs167771), Dopamine receptor D4 (*DRD4*; rs1800955), Gamma-aminobutyric acid type A receptor subunit alpha6 (*GABRA6*; rs3219151), and Oxytocin receptor (*OXTR*; rs2254295) (see **Table 7.1**). These gene names and symbols are in accordance with those officially approved by the Human Gene Nomenclature Committee (HGNC; https://www.genenames.org). Standard genomic quality control procedures and thresholds were applied when selecting polymorphisms: SNP call rate (>95), sample call rate (>95), fisher's linear discriminant (>3.6), and minor allele frequency (>0.05).

Table 7.1. Psychological gene and single nucleotide polymorphism (SNP) information.

Gene	Symbol	Chr	Function	SNP	Consequence	Associations	MAF
5-hydroxytryptamine receptor 2A	*HTR2A*	13q14.2	Encodes one of the receptors for serotonin, a neurotransmitter with an important role in the brain reward system as well as cellular development and differentiation.	rs6311	Intron variant C > T	Big Five personality traits and novelty seeking.	T = 0.44
Brain derived neurotrophic factor	*BDNF*	11p14.1	Encodes a neurotrophin essential for neuronal development and survival, synaptic plasticity, and cognitive function.	rs6265	Missense variant C > T (Val > Met)	Novelty seeking and sensation seeking.	T = 0.20
Cholinergic receptor muscarinic 2	*CHRM2*	7q33	Muscarinic receptors influence many effects of acetylcholine in the central and peripheral nervous system. The muscarinic cholinergic receptor 2 is involved in mediation of bradycardia and a decrease in cardiac contractility	rs1824024	Intron variant C > A	Big Five personality traits and sensation seeking.	C = 0.29
Catechol-O-methyltransferase	*COMT*	22q11.21	Catalyses the degradation of catecholamines, including neurotransmitters dopamine, epinephrine, and norepinephrine.	rs4680	Missense variant G > A (Val > Met)	Impulsivity, addiction disorders, extraversion, sensation seeking.	A = 0.50
Catenin alpha 2	*CTNNA2*	2p12	Expressed in the brain with a role in synaptic plasticity, actin filament binding activity, regulation of Arp2/3 complex-mediated actin nucleation, neuron migration and projection development.	rs7600563	Intron variant T > G	Excitement seeking and risk taking.	G = 0.34

Dopamine receptor D2	*DRD2*	11q23.2	Encodes the D2 subtype of the dopamine receptor, which inhibits adenylyl cyclase activity and regulates neuronal growth, development, and behaviour.	rs1800497	Missense variant G > A (Glu > Lys)	Sensation seeking, impulsive behaviour, and novelty seeking.	A = 0.19
Dopamine receptor D3	*DRD3*	3q13.31	Encodes the D3 subtype of the dopamine receptor, which inhibits adenylyl cyclase activity and is associated with cognitive, emotional, and endocrine functions.	rs167771	Intron variant G > A	Sensation seeking and autism spectrum disorders.	G = 0.20
Dopamine receptor D4	*DRD4*	11p15.5	Encodes the D4 subtype of the dopamine receptor, which inhibits adenylyl cyclase activity and is associated with autonomic nervous system dysfunction, behaviour and personality traits.	rs1800955	2KB Upstream variant T > C	Novelty seeking, sensation seeking, anxiety and impulsivity.	C = 0.41
Gamma-aminobutyric acid type A receptor subunit alpha6	*GABRA6*	5q34	GABA is the major inhibitory neurotransmitter in the mammalian brain where it acts at GABA-A receptors, which are ligand-gated chloride channels.	rs3219151	3 Prime UTR variant C > T	Big Five personality traits, panic and major depressive disorder.	C = 0.42
Oxytocin receptor	*OXTR*	3p25.3	Encodes oxytocin a neuropeptide and hormone implicated in prosocial human social and emotional functioning.	rs2254295	Intron variant C > T	Reward seeking, risk taking, and impulsivity.	C = 0.11

Note. Chr = chromosome location; MAF = minor allele frequency (according to European population; 1000 Genomes Project Consortium, 2015).

7.2.4.3 Total genotype score

Unweighted and weighted total genotype scores (TGS; TWGS) were calculated to assess the combined influence of the included SNPs on each personality dimension and mental toughness. Both TGSs and TWGSs have demonstrated sufficient discriminatory power in previous sport genomic research (Massidda et al., 2014; Varillas Delgado et al., 2020; Williams & Folland, 2008). To generate both the TGS and TWGS, each genotype of a respective SNP initially received a score between 0-2 based on the observed associations with a dependent variable. Genotypes of dominant (AA vs. Aa-aa) and recessive (AA-Aa vs. aa) models were assigned a score of two (i.e., associated genotype[s]) or zero (i.e., alternate genotype[s]), whereas genotypes of co-dominant models (AA vs. Aa vs. aa) were assigned three scores (i.e., homozygous-associated genotypes received a score of two, the heterozygote received a score of one, and the alternate homozygous genotype received a score of zero) (Guilherme et al., 2020).

For the TGS, the genotype scores (GS) were then summed and transformed into a 0-100 scale by dividing the total score by the maximum possible score and multiplying by 100.

$$TGS = (combined\text{-}GS / maximum\text{-}GS) * 100$$

For the TWGS, a similar procedure to Varillas Delgado et al. (2020) was used. Each GS was multiplied by the standardised beta coefficients (β) of each SNP following multiple regression with each dependent variable to create weighted genotype scores (WGS). The WGSs were then summed and transformed into a 0-100 scale by dividing the total score by the maximum possible score and multiplying by 100.

$$TWGS = (combined\text{-}WGS / maximum\text{-}WGS) * 100$$

7.2.5 Data analysis

Each SNP was tested for adherence with Hardy-Weinberg equilibrium (HWE) using an exact test via SNPStats (Solé et al., 2006). Linkage disequilibrium (LD) was analysed using LDlink (Machiela & Chanock, 2015) and data from the 1000 Genomes Project European ancestry population (1000 Genomes Project Consortium, 2015). All other data were analysed using Jamovi version 1.8.1 and IBM SPSS version 25. Normality was assessed with the Shapiro-Wilk test and homoscedasticity was assessed using Levene's test. Akaike information criterion (AIC) was used to select which genetic model (i.e., co-dominant, dominant, recessive) best fit the data and would be subjected to hypothesis testing. However, if MAF ≤ 0.25 a dominant model was utilised to retain statistical power (Murtagh et al., 2020). Simple linear regression was performed to assess the association of genotype models with each personality dimension and mental toughness. Multiple regression was used to calculate the standardised beta coefficients (β) of each SNP for the TWGS models. Simple linear regression was then performed to assess the association of each TGS and TWGS with each personality dimension and mental toughness. Pearson's correlation coefficient (*r*) with thresholds values of ≤0.1 (trivial), >0.1-0.3 (small), >0.3–0.5 (moderate), >0.5–0.7 (large), >0.7–0.9 (very large), and >0.9–1.0 (almost perfect) were used to measure correlation (Hopkins, 2009). The coefficient of determination (R^2) was computed to determine the variance explained by each TGS and TWGS. Statistical significance was set at $p < .05$.

7.3 Results

Genotype and allele distributions of all SNPs were in HWE except for *DRD4* ($p = .040$), and all SNPs were in linkage equilibrium. Assumptions of normality and homoscedasticity were not violated. Descriptive statistics and genotype frequencies are displayed in **Table 7.2**.

Table 7.2. Descriptive statistics of personality dimensions and mental toughness.

Gene (SNP)	Genotype = *n* (%)	O	A	C	N	E	MT	MAF	HWE
Overall		25.47 ± 5.46	26.27 ± 4.98	27.67 ± 5.07	22.78 ± 6.85	23.67 ± 6.75	43.21 ± 5.56	N/A	N/A
	C/C = 28 (38)	25.96 ± 4.99	25.18 ± 4.59	27.71 ± 4.04	23.79 ± 7.41	25.39 ± 6.41	43.46 ± 5.57		
HTR2A (rs6311)	C/T = 29 (40)	25.10 ± 5.17	27.59 ± 5.58	27.69 ± 5.86	22.03 ± 6.11	21.03 ± 7.21	42.63 ± 5.60	0.42	0.15
	T/T = 16 (22)	25.25 ± 6.88	25.81 ± 4.17	27.56 ± 5.48	22.38 ± 7.31	25.44 ± 5.02	43.75 ± 5.76		
	C/C = 45 (62)	25.40 ± 5.55	26.91 ± 4.54	27.82 ± 5.58	23.47 ± 7.31	24.16 ± 6.36	43.53 ± 5.29		
BDNF (rs6265)	C/T = 25 (34)	26.20 ± 5.34	25.68 ± 5.67	27.36 ± 4.35	21.16 ± 6.12	22.32 ± 7.39	42.25 ± 6.28	0.21	1
	T/T = 3 (4)	20.33 ± 2.31	21.67 ± 3.06	28.00 ± 3.61	26.00 ± 2.00	27.67 ± 6.51	45.50 ± 3.32		
	A/A = 23 (32)	26.22 ± 5.97	27.09 ± 5.83	27.87 ± 5.55	20.83 ± 8.01	24.04 ± 4.88	42.48 ± 6.80		
CHRM2 (rs1824024)	A/C = 41 (56)	24.90 ± 5.37	25.73 ± 4.48	27.44 ± 4.78	23.78 ± 5.84	23.54 ± 7.50	43.97 ± 4.32	0.40	0.22
	C/C = 9 (12)	26.11 ± 4.73	26.67 ± 5.07	28.22 ± 5.63	23.22 ± 7.66	23.33 ± 7.91	41.78 ± 6.94		
	A/A = 13 (18)	26.31 ± 5.02	24.92 ± 5.02	27.69 ± 4.61	19.54 ± 5.95	21.46 ± 5.61	44.85 ± 6.39		
COMT (rs4680)	G/A = 34 (47)	25.00 ± 5.05	26.85 ± 5.08	27.41 ± 5.31	23.71 ± 6.97	24.59 ± 6.58	42.38 ± 4.28	0.41	0.81
	G/G = 26 (36)	25.65 ± 6.27	26.19 ± 4.88	28.00 ± 5.15	23.19 ± 6.86	23.58 ± 7.43	43.42 ± 6.48		
	G/G = 3 (4)	24.33 ± 9.50	24.67 ± 4.16	27.33 ± 8.50	26.00 ± 7.55	28.00 ± 9.64	44.33 ± 5.13		
CTNNA2 (rs7600563)	T/G = 29 (40)	25.90 ± 6.20	25.72 ± 4.83	27.86 ± 4.44	22.69 ± 7.84	23.03 ± 6.69	43.89 ± 5.86	0.24	0.54
	T/T = 40 (56)	25.20 ± 4.73	26.60 ± 5.11	27.30 ± 5.17	22.70 ± 6.19	23.88 ± 6.72	42.56 ± 5.49		
DRD2 (rs1800497)	A/A = 5 (7)	25.20 ± 5.40	24.00 ± 4.24	26.60 ± 2.30	25.20 ± 5.89	24.80 ± 5.89	40.60 ± 3.65	0.25	0.75
	G/A = 26 (36)	25.69 ± 5.55	26.54 ± 5.62	27.35 ± 5.61	22.38 ± 6.31	23.77 ± 7.34	42.91 ± 4.43		

	G/G = 42 (58)	25.36 ± 5.53	26.38 ± 4.68	28.00 ± 5.02	22.74 ± 7.34	23.48 ± 6.60	43.67 ± 6.24		
	A/A = 46 (63)	25.41 ± 6.07	27.17 ± 5.30	28.15 ± 5.09	23.00 ± 7.37	24.07 ± 7.52	43.20 ± 5.98		
DRD3 (rs167771)	A/G = 23 (32)	25.39 ± 4.25	24.87 ± 3.96	27.26 ± 4.96	22.30 ± 6.31	23.65 ± 5.12	43.81 ± 4.98	0.21	0.72
	G/G = 4 (5)	26.50 ± 5.26	24.00 ± 5.03	24.50 ± 5.51	23.00 ± 3.92	19.25 ± 4.86	40.25 ± 2.63		
	C/C = 18 (25)	25.78 ± 4.67	27.06 ± 5.77	26.94 ± 5.00	20.39 ± 7.92	22.06 ± 6.65	41.94 ± 7.30		
DRD4 (rs1800955)	T/C = 27 (37)	25.56 ± 6.25	27.04 ± 4.90	28.19 ± 5.56	23.22 ± 6.41	25.11 ± 6.51	44.21 ± 4.74	0.43	0.04
	T/T = 28 (38)	25.18 ± 5.28	25.04 ± 4.41	27.64 ± 4.74	23.89 ± 6.37	23.32 ± 6.99	43.13 ± 5.07		
	C/C = 11 (15)	24.64 ± 3.91	26.73 ± 4.50	29.27 ± 5.57	21.82 ± 6.68	25.00 ± 5.46	42.83 ± 5.13		
GABRA6 (rs3219151)	T/C = 41 (58)	25.37 ± 5.75	26.49 ± 5.22	27.29 ± 4.52	23.07 ± 6.40	23.32 ± 6.32	43.53 ± 5.39	0.44	0.23
	T/T = 19 (27)	26.00 ± 6.01	24.95 ± 4.62	27.68 ± 4.84	23.47 ± 7.95	23.37 ± 8.43	42.70 ± 6.42		
	C/C = 3 (4)	21.67 ± 5.03	27.67 ± 5.51	29.00 ± 8.19	21.00 ± 4.36	21.00 ± 6.00	47.33 ± 5.13		
OXTR (rs2254295)	T/C = 17 (23)	26.12 ± 5.06	26.12 ± 4.92	27.65 ± 4.36	23.65 ± 5.99	23.71 ± 7.50	41.13 ± 4.63	0.16	0.36
	T/T = 53 (73)	25.47 ± 5.60	26.25 ± 5.06	27.60 ± 5.20	22.60 ± 7.26	23.81 ± 6.63	43.62 ± 5.71		

Note. Data presented in mean ± standard deviation. O = openness; A = agreeableness; C = conscientiousness; N = neuroticism; E = extraversion; MT = mental toughness; MAF = minor allele frequency; HWE = Hardy-Weinberg equilibrium.

7.3.1 Individual SNPs

There was a significant association between *DRD3* ($F_{(1, 71)} = 4.24$, $p = .043$) and agreeableness, where A/A homozygotes scored higher than G allele carriers ($B = 2.43$). No other associations were found (see **Table 7.3**).

7.3.2 TGS

Associations were also noted between the TGS and agreeableness ($F_{(1, 69)} = 6.31$, $p = .014$), extraversion ($F_{(1, 69)} = 7.16$, $p = .009$), mental toughness ($F_{(1, 68)} = 5.77$, $p = .019$), and neuroticism ($F_{(1, 69)} = 8.40$, $p = .005$). Moreover, small positive correlations were found with agreeableness ($r = .29$; $R^2 = .08$) and mental toughness ($r = .28$; $R^2 = .08$), whilst moderate positive correlations were found with extraversion ($r = .31$; $R^2 = .09$) and neuroticism ($r = .33$; $R^2 = .11$) (see **Figure 7.1**, **Figure 7.2**, and **Figure 7.3**).

7.3.3 TWGS

There were significant associations between the TWGS and agreeableness ($F_{(1, 69)} = 10.57$, $p = .002$), conscientiousness ($F_{(1, 69)} = 4.30$, $p = .042$), extraversion ($F_{(1, 69)} = 9.50$, $p = .003$), mental toughness ($F_{(1, 68)} = 7.85$, $p = .007$), and neuroticism ($F_{(1, 69)} = 14.58$, $p < .001$). Moreover, a small positive correlation was found with conscientiousness ($r = .24$; $R^2 = .06$), whilst moderate positive correlations were found with agreeableness ($r = .35$; $R^2 = .12$), extraversion ($r = .36$; $R^2 = .13$), mental toughness ($r = .32$; $R^2 = .10$), and neuroticism ($r = .42$; $R^2 = .17$) (see **Figure 7.4**, **Figure 7.5**, and **Figure 7.6**).

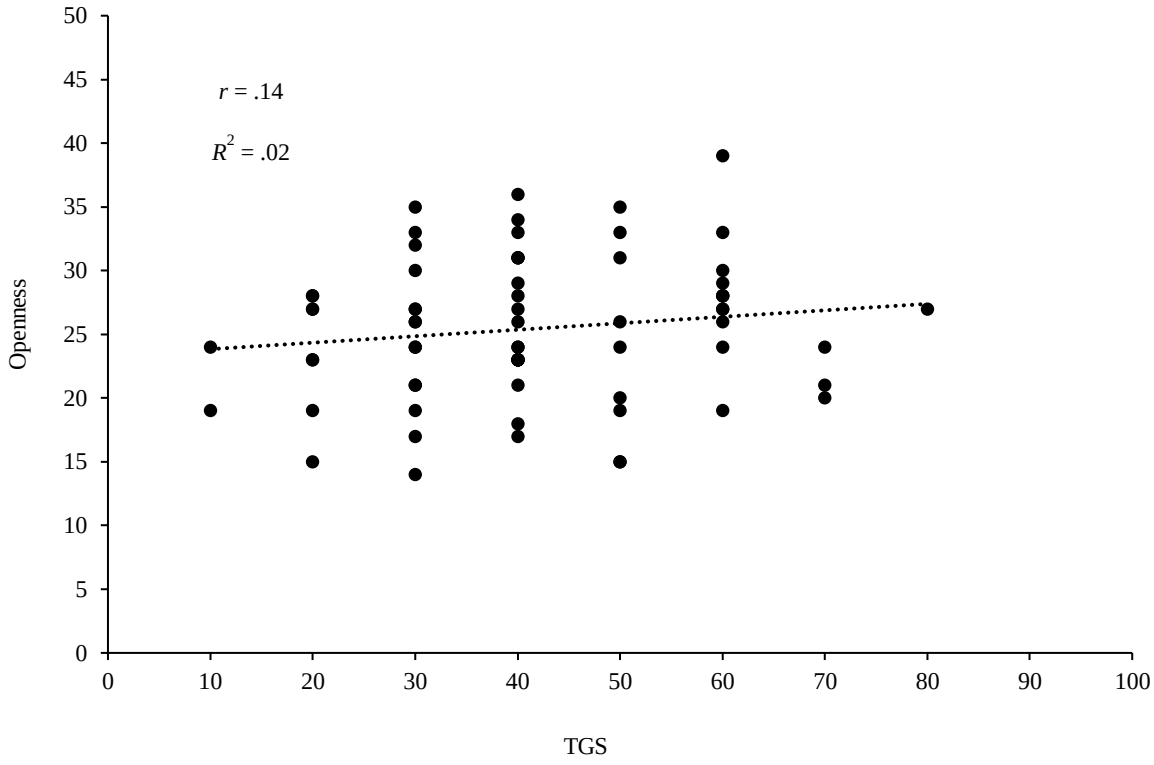
Table 7.3. Psychological simple linear regression analysis.

Gene (SNP)	Model	*B*	*SE B*	β	*t*	*p*
Openness						
HTR2A (rs6311)	T/T-T/C vs. C/C	-0.81	1.32	-0.15	-0.61	.542
BDNF (rs6265)	T/T-T/C vs. C/C	0.17	1.32	0.03	0.13	.897
CHRM2 (rs1824024)	C/C-C/A vs. A/A	-1.10	1.38	-0.20	-0.80	.429
COMT (rs4680)	G/G-G/A vs. A/A	-1.02	1.68	-0.19	-0.61	.543
CTNNA2 (rs7600563)	T/T vs. T/G-G/G	-0.55	1.31	-0.10	-0.42	.676
DRD2 (rs1800497)	G/G vs. G/A-A/A	-0.26	1.30	-0.05	-0.20	.845
DRD3 (rs167771)	G/G-G/A vs. A/A	0.14	1.33	0.03	0.11	.915
DRD4 (rs1800955)	T/T vs. T/C-C/C	-0.47	1.32	-0.09	-0.35	.726
GABRA6 (rs3219151)	T/T vs. T/C-C/C	0.79	1.49	0.14	0.53	.598
OXTR (rs2254295)	T/T vs. T/C-C/C	0.02	1.44	0.00	0.02	.988
Agreeableness						
HTR2A (rs6311)	T/T-T/C vs. C/C	1.78	1.19	0.36	1.49	.139
BDNF (rs6265)	T/T-T/C vs. C/C	-1.66	1.19	-0.33	-1.39	.168
CHRM2 (rs1824024)	C/C-C/A vs. A/A	-1.19	1.26	-0.24	-0.95	.348
COMT (rs4680)	G/G-G/A vs. A/A	1.64	1.52	0.33	1.08	.284
CTNNA2 (rs7600563)	T/T vs. T/G-G/G	0.98	1.17	0.20	0.83	.408
DRD2 (rs1800497)	G/G vs. G/A-A/A	0.25	1.19	0.05	0.21	.833
DRD3 (rs167771)	G/G-G/A vs. A/A	-2.43	1.18	-0.49	-2.06	**.043***
DRD4 (rs1800955)	T/T vs. T/C-C/C	-2.01	1.18	-0.40	-1.70	.094
GABRA6 (rs3219151)	T/T vs. T/C-C/C	-1.59	1.32	-0.32	-1.20	.233

OXTR (rs2254295)	T/T vs. T/C-C/C	-0.10	1.32	-0.02	-0.08	.937
Conscientiousness						
HTR2A (rs6311)	T/T vs. T/C-C/C	-0.14	1.44	-0.03	-0.10	.923
BDNF (rs6265)	T/T-T/C vs. C/C	-0.39	1.23	-0.08	-0.32	.750
CHRM2 (rs1824024)	C/C vs. C/A-A/A	0.63	1.82	0.12	0.35	.730
COMT (rs4680)	G/G vs. G/A-A/A	0.51	1.25	0.10	0.41	.683
CTNNA2 (rs7600563)	T/T vs. T/G-G/G	-0.51	1.18	-0.10	-0.43	.666
DRD2 (rs1800497)	G/G vs. G/A-A/A	0.77	1.21	0.15	0.64	.523
DRD3 (rs167771)	G/G-G/A vs. A/A	-1.30	1.23	-0.26	-1.06	.293
DRD4 (rs1800955)	T/T-T/C vs. C/C	0.96	1.38	0.19	0.70	.488
GABRA6 (rs3219151)	T/T-T/C vs. C/C	-1.86	1.56	-0.39	-1.19	.237
OXTR (rs2254295)	T/T vs. T/C-C/C	-0.25	1.34	-0.05	-0.18	.855
Neuroticism						
HTR2A (rs6311)	T/T-T/C vs. C/C	-1.63	1.65	-0.24	-0.99	.326
BDNF (rs6265)	T/T-T/C vs. C/C	-1.79	1.65	-0.26	-1.09	.281
CHRM2 (rs1824024)	C/C-C/A vs. A/A	2.85	1.70	0.42	1.68	.098
COMT (rs4680)	G/G-G/A vs. A/A	3.94	2.06	0.58	1.92	.059
CTNNA2 (rs7600563)	T/T vs. T/G-G/G	-0.30	1.64	-0.04	-0.18	.856
DRD2 (rs1800497)	G/G vs. G/A-A/A	-0.10	1.63	-0.01	-0.06	.951
DRD3 (rs167771)	G/G-G/A vs. A/A	-0.59	1.67	-0.09	-0.35	.724
DRD4 (rs1800955)	T/T-T/C vs. C/C	3.17	1.83	0.46	1.73	.088
GABRA6 (rs3219151)	T/T-T/C vs. C/C	1.38	2.24	0.20	0.62	.540
OXTR (rs2254295)	T/T vs. T/C-C/C	-0.65	1.81	-0.09	-0.36	.722

Extraversion						
HTR2A (rs6311)	T/T-T/C vs. C/C	-2.79	1.60	-0.41	-1.74	.086
BDNF (rs6265)	T/T-T/C vs. C/C	-1.26	1.63	-0.19	-0.78	.441
CHRM2 (rs1824024)	C/C-C/A vs. A/A	-0.54	1.71	-0.08	-0.32	.752
COMT (rs4680)	G/G-G/A vs. A/A	2.69	2.05	0.40	1.31	.195
CTNNA2 (rs7600563)	T/T vs. T/G-G/G	0.37	1.62	0.06	0.23	.818
DRD2 (rs1800497)	G/G vs. G/A-A/A	-0.46	1.61	-0.07	-0.29	.776
DRD3 (rs167771)	G/G-G/A vs. A/A	-1.07	1.64	-0.16	-0.65	.519
DRD4 (rs1800955)	T/T-T/C vs. C/C	2.14	1.83	0.32	1.17	.245
GABRA6 (rs3219151)	T/T-T/C vs. C/C	-1.67	2.22	-0.25	-0.75	.456
OXTR (rs2254295)	T/T vs. T/C-C/C	0.51	1.78	0.08	0.29	.775
Mental toughness						
HTR2A (rs6311)	T/T vs. T/C-C/C	0.70	1.59	0.13	0.44	.663
BDNF (rs6265)	C/C vs. C/T-T/T	0.82	1.36	0.15	0.60	.547
CHRM2 (rs1824024)	A/A vs. A/C-C/C	-1.08	1.41	-0.19	-0.77	.446
COMT (rs4680)	G/G-G/A vs. A/A	-2.00	1.70	-0.36	-1.18	.244
CTNNA2 (rs7600563)	G/G-G/T vs. T/T	1.37	1.35	0.25	1.02	.312
DRD2 (rs1800497)	A/A-A/G vs. G/G	-1.17	1.35	-0.21	-0.87	.388
DRD3 (rs167771)	A/A vs. A/G-G/G	-0.04	1.39	-0.01	-0.03	.975
DRD4 (rs1800955)	T/T-T/C vs. C/C	1.67	1.55	0.30	1.08	.284
GABRA6 (rs3219151)	T/T vs. T/C-C/C	-0.66	1.49	-0.12	-0.44	.659
OXTR (rs2254295)	C/C-C/T vs. T/T	-1.51	1.49	-0.27	-1.01	.315

Note. Bold values and * highlight statistical significance at $p < .05$. *B* = unstandardised beta; *SE B* = standard error; β = standardised beta.



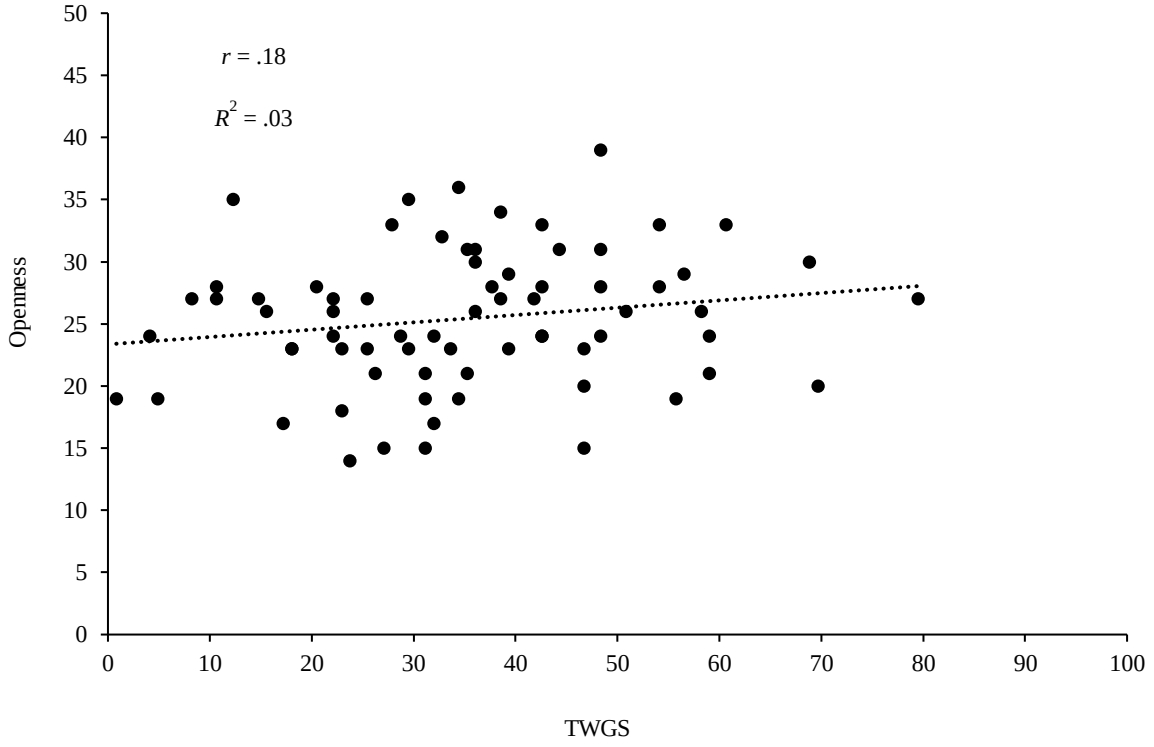


Figure 7.1. Openness total genotype score (TGS) and total weighted genotype score (TWGS) correlations. * Statistically significant at $p < .05$.

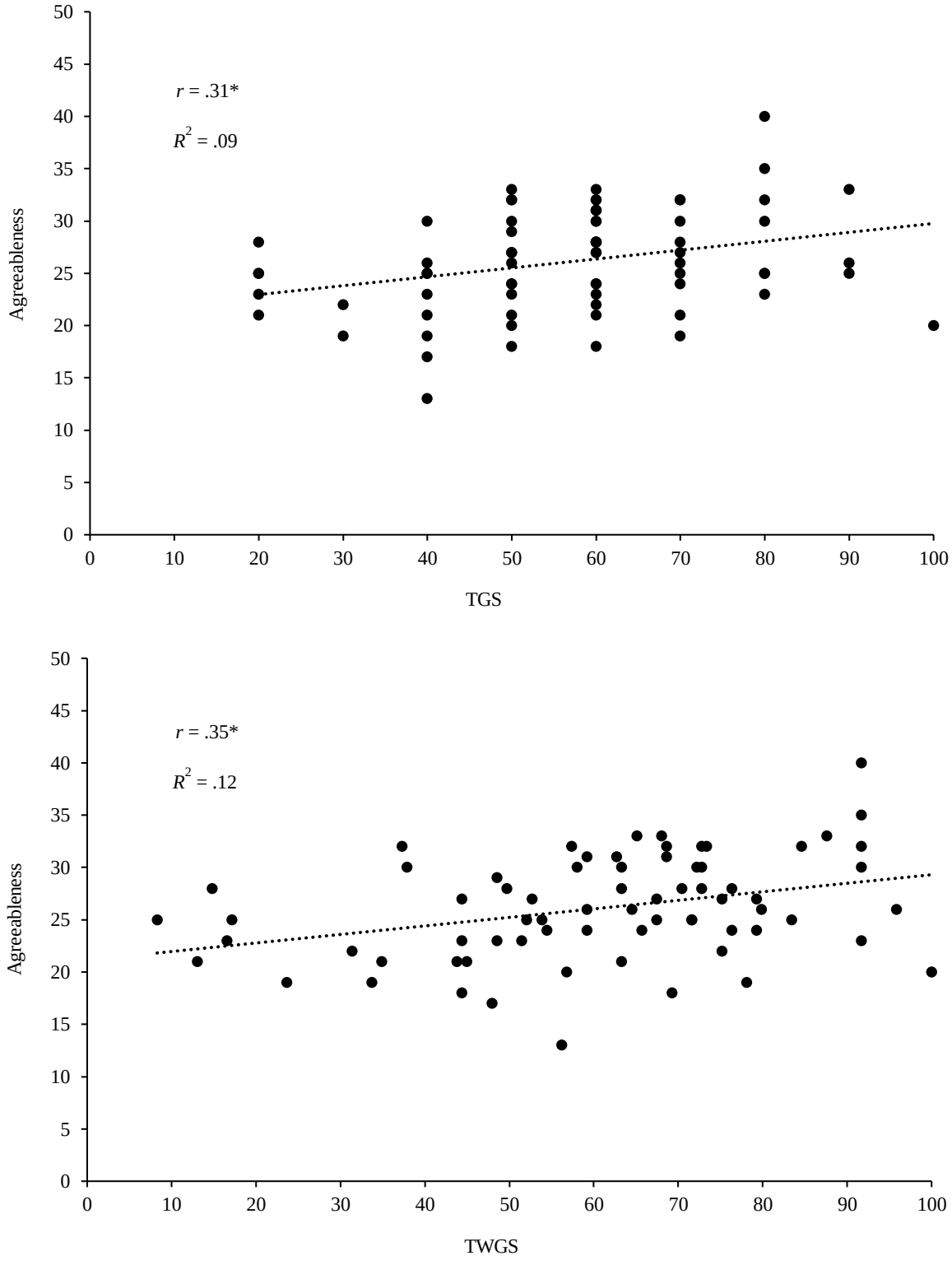


Figure 7.2. Agreeableness total genotype score (TGS) and total weighted genotype score (TWGS) correlations. * Statistically significant at $p < .05$.

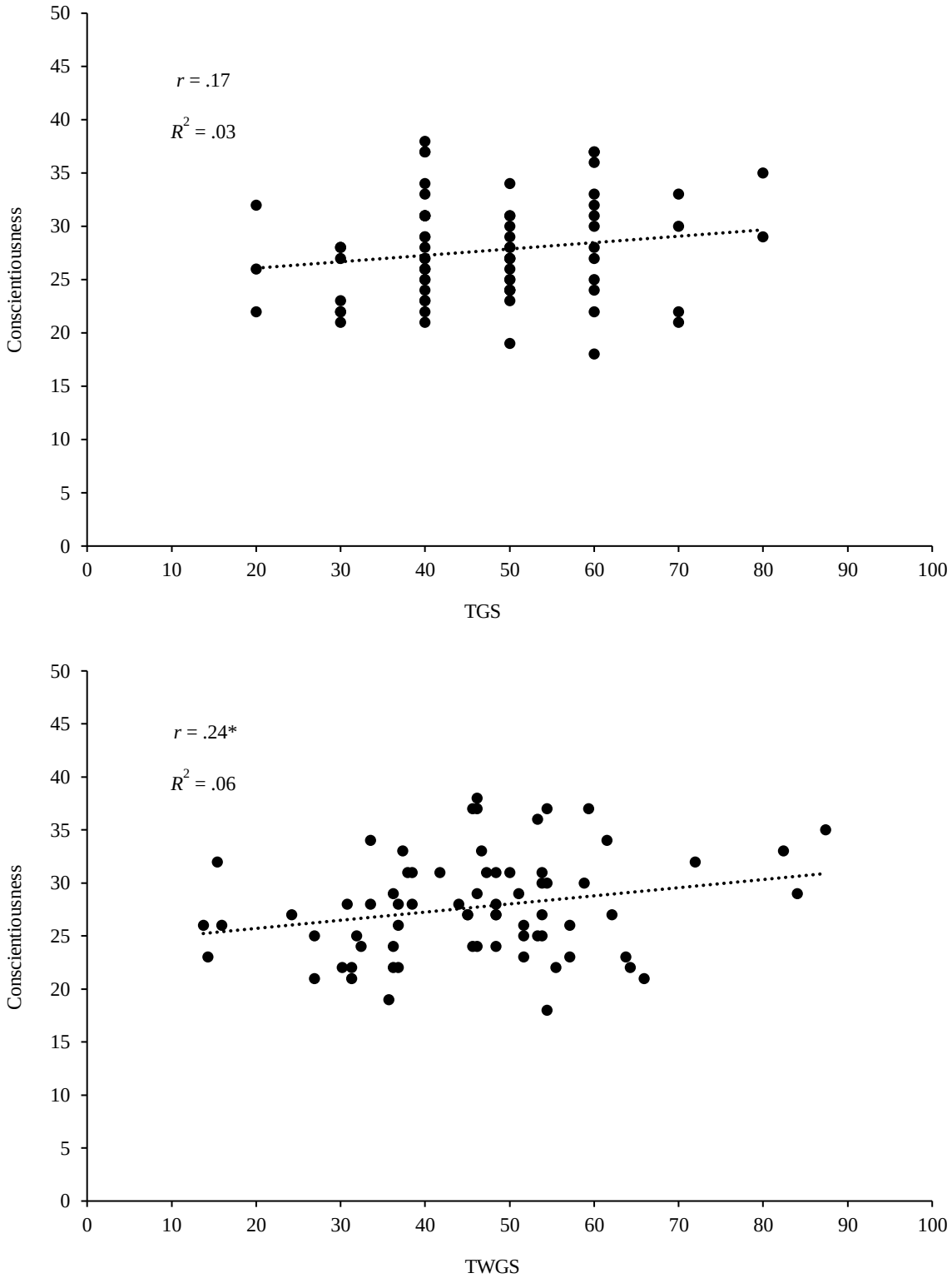


Figure 7.3. Conscientiousness total genotype score (TGS) and total weighted genotype score (TWGS) correlations. * Statistically significant at $p < .05$.

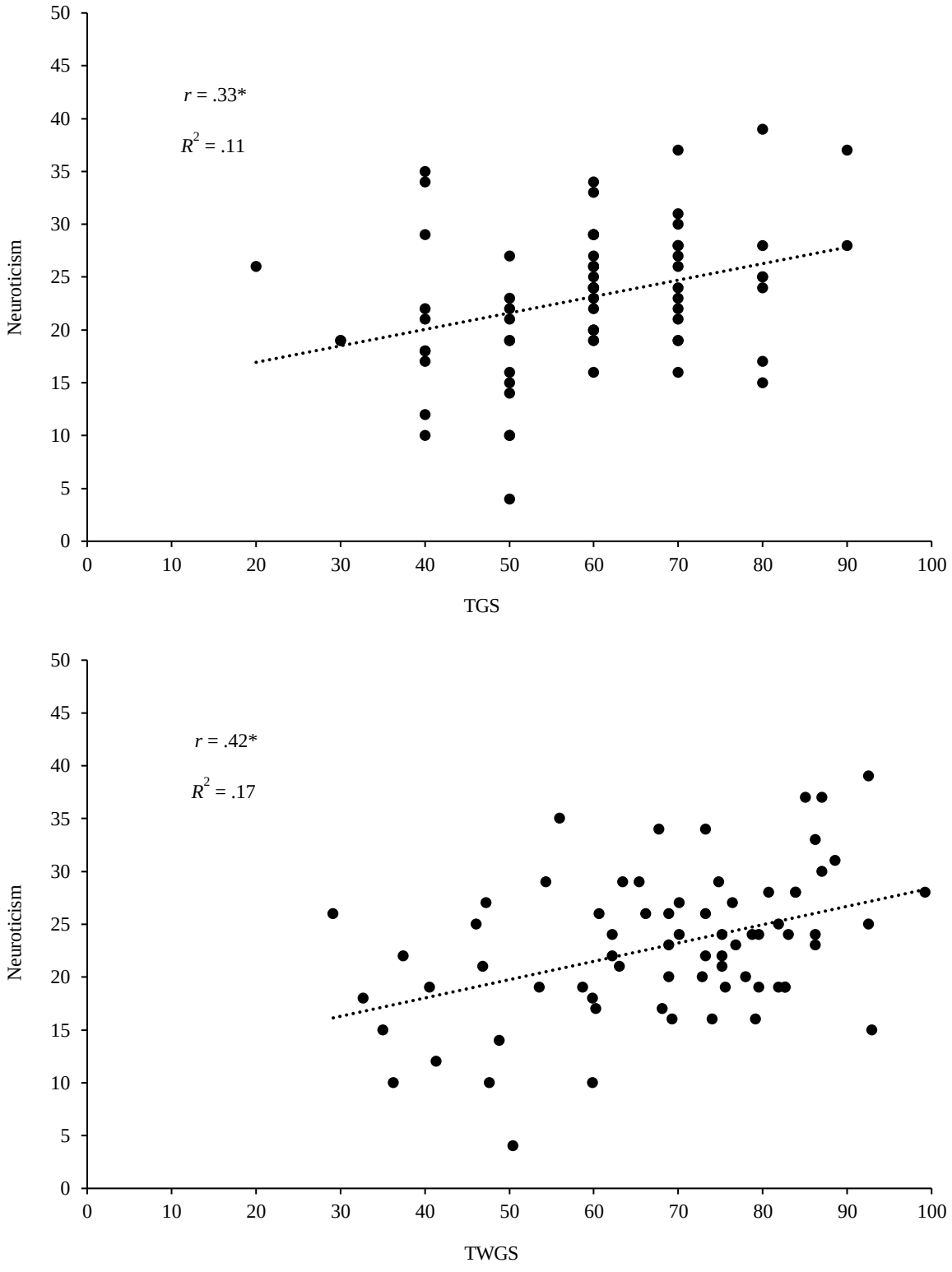


Figure 7.4. Neuroticism total genotype score (TGS) and total weighted genotype score (TWGS) correlations. * Statistically significant at $p < .05$.

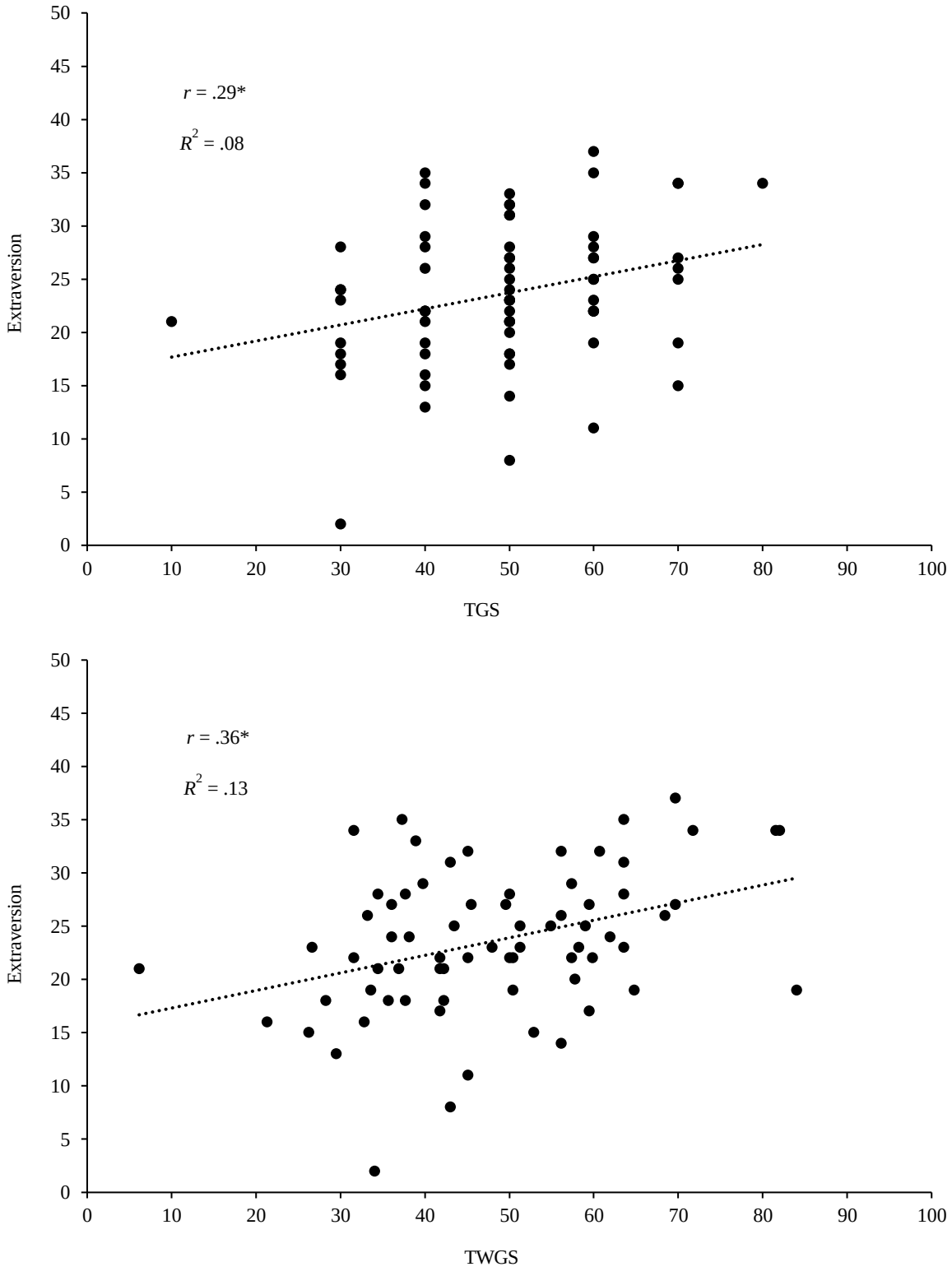


Figure 7.5. Extraversion total genotype score (TGS) and total weighted genotype score (TWGS) correlations. * Statistically significant at $p < .05$.

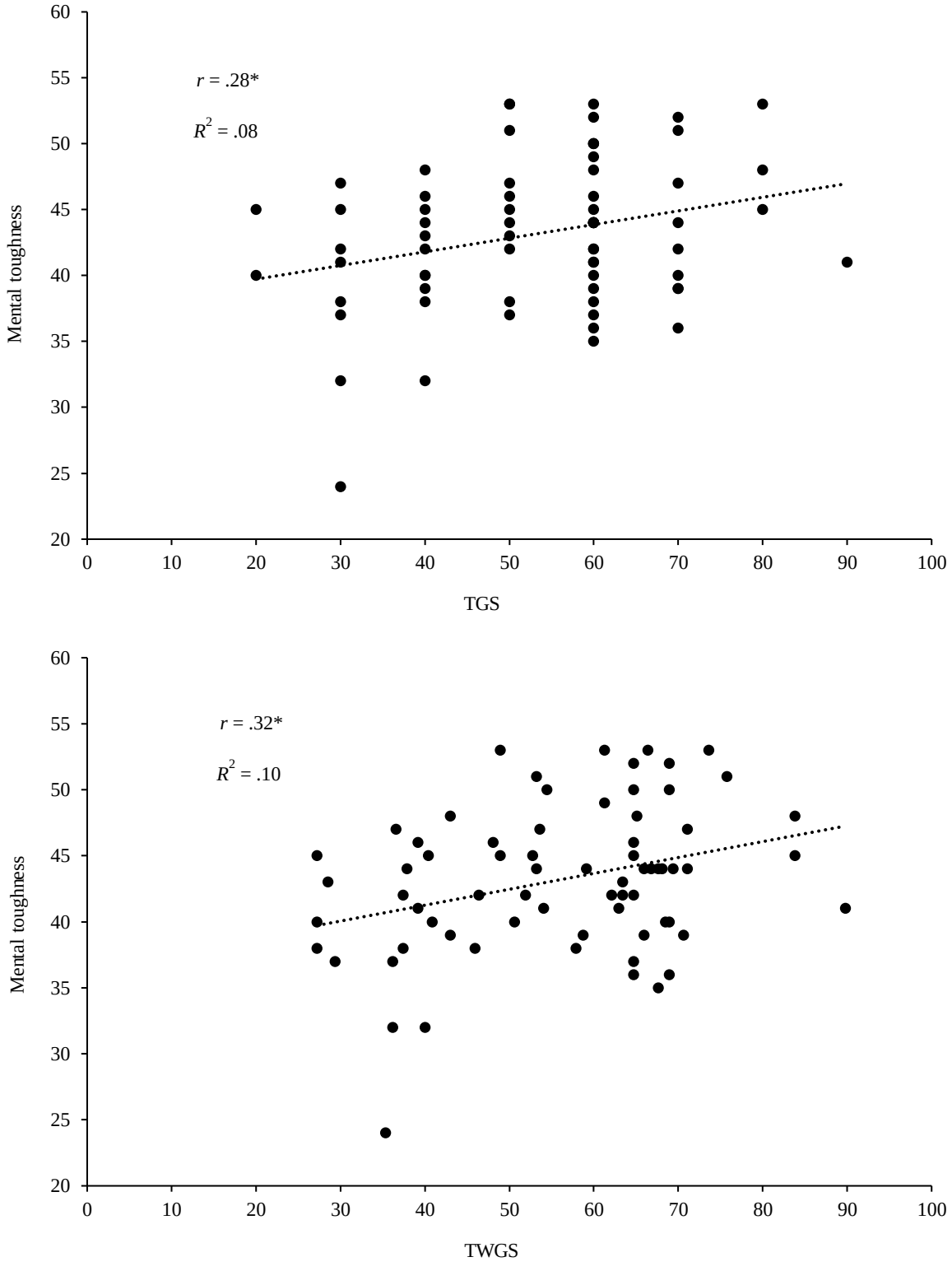


Figure 7.6. Mental toughness total genotype score (TGS) and total weighted genotype score (TWGS) correlations. * Statistically significant at $p < .05$.

7.4 Discussion

This study examined associations amongst ten psychogenetic polymorphisms, both individually and collectively, with the personality and mental toughness profiles of academy football players. To the author's knowledge, this is the first assessment of the influence of genetic markers in isolation, and as part of a polygenic profile, on psychological traits within a homogenous football cohort. This study presents a novel association of *DRD3* (rs167771) with agreeableness and a preliminary polygenic model that explains a small proportion of the inter-individual variance in mental toughness and each personality dimension except openness. As such, these findings suggest several psychological characteristics of youth football players may be partly determined by genetic factors.

The A/A genotype of *DRD3* (rs167771) was associated with higher levels of agreeableness compared to the G allele. The *DRD3* gene encodes the DRD3 protein, which is the D3 subtype of the five dopamine receptors highly expressed in the limbic regions of the brain (e.g., hippocampus, nucleus accumbens, and ventral striatum) (Bouthenet et al., 1991; Gurevich & Joyce, 1999). The *DRD3* (rs167771) SNP is an intron (i.e., non-coding elements of a gene), which are important for efficient splicing or other regulatory elements involved in transcription (Kim et al., 2008). In line with the present study, *DRD3* has previously been associated with emotional and cognitive functions (Bouthenet et al., 1991; Gurevich & Joyce, 1999). However, at the time of writing, there is no data available regarding the underpinning mechanism of the allelic association between *DRD3* (rs167771) and psychological traits, so it remains unclear whether they have a functional effect or are in LD with the actual effect alleles (Staal, 2015).

To the authors' knowledge, this is the first study to find a direct association between *DRD3* (rs167771) and agreeableness. This may be due to the unique participants under investigation in this study. However, this finding is comparable to previous relationships between *DRD3* (rs167771) and other relevant phenotypes. For instance, previous research has indicated that the G allele of *DRD3* (rs167771) may be associated with an increased risk of autism (de Krom et al., 2009; Toma et al., 2013). Autism is an applicable phenotype to use for an indirect comparison, as lower levels of agreeableness have also been associated with an increased risk of autism (Austin, 2005). Although further research is required, these findings are consistent with the scientific literature associating *DRD3* (rs167771) with psychological traits.

There was no association between any other SNP in isolation and a personality dimension or mental toughness. Accurately measuring personality dimensions and mental toughness is notoriously difficult. Moreover, their overall phenotypic complexity means it is highly unlikely that inter-individual variance would be explained by individual SNPs. Indeed, identifying genetic associations between common variants and psychological traits may be even more challenging than with physical traits. For instance, whilst the SNPs associated with the largest effects on physical traits explain an estimated 0.3% of the inter-individual variance, the SNPs associated with the largest effects on psychological traits only explain an estimated 0.02% of the inter-individual variance (Chabris et al., 2015). This suggests a larger number of SNPs with smaller effect sizes may account for the variability observed in psychological traits. Moreover, it was shown that a psychological condition (i.e., schizophrenia) was influenced by ~5,000 more SNPs than four physical conditions (i.e., celiac disease, coronary artery disease, rheumatoid arthritis, and type 2 diabetes) (Ripke et al., 2013).

Candidate gene approaches are exemplified by numerous limitations in psychological research (see Munafò & Flint, 2011), but can also confer inferential advantages when sample cohorts are small and/or unique (e.g., athletic populations) (Jorgensen et al., 2009). However, evidence suggests a vast number of SNPs with small effect sizes may underpin psychological phenotypes. In light of this, polygenic profiles were used to assess the combined influence of the selected SNPs, as they have not only proved effective in psychogenetic research (Ausmees et al., 2021), but also previous research on phenotypes in football (e.g., athlete status, injury, physiological performance, and technical ability) (McAuley, Hughes, et al., 2021a). Moreover, applying a weight to each SNP based on its individual influence on a respective phenotype has been shown to increase the variance explained by a polygenic profile on physiological and technical phenotypes in football (Massidda et al., 2014). The results from this study correspond with and expand previous research, as the TGSs and TWGSs were associated with several psychological phenotypes, but the TWGS consistently explained more variance.

The polygenic profiles in this study were associated with mental toughness and all personality dimensions except openness. The combined variance explained by the SNPs in the polygenic models differed between each psychological phenotype, where the strongest association was observed with neuroticism (e.g., neuroticism = 11-17%, extraversion = 9-13%, agreeableness = 8-12%, mental toughness = 8-10%, conscientiousness = 3-6%, and openness = 2-3%). All personality dimensions and mental toughness have been previously associated with differentiating successful and less successful individual and team-sport athletes (Benítez-Sillero et al., 2021; Piepiora, 2021; Steca et al., 2018). That said, more consistent associations and larger effect sizes are generally reported with neuroticism, which is encompassed by facets such as anxiety, depression, hostility, impulsiveness, self-consciousness, and vulnerability

(Piepiora & Piepiora, 2021). As such, the most noteworthy feature of this study is the identification of a panel of SNPs that may be associated with one of the key psychological variables in sport, and more specifically football.

These SNPs may prove useful in creating a genetic tool capable of assisting practitioners with implementing more individualised psychological intervention programmes in the future. That said, precisely how genetic information should be utilised by practitioners is still unclear and requires careful thought and consideration (McAuley, Hughes, et al., 2021b). Reducing the complex biological nature of psychological phenotypes to answers in a questionnaire is ill-advised. The intricacies of an individuals' moment to moment experiences can be difficult to articulate, perceptual, and difficult to measure, even if more robust measures are used. Therefore, the validation of these results in larger homogenous and independent football cohorts is required, as well as the identification of more relevant genetic variants, before implementing the current findings into practice. However, even if these results are validated in larger independent cohorts, we believe they should not be used for talent identification purposes due to the multifactorial nature of performance in football and the accompanying social, ethical, and legal issues associated with potential genetic discrimination (McAuley, Hughes, et al., 2022). Instead, validated results could be used to enhance athlete development processes by managing athlete welfare to facilitate improved long-term performance. Genetic information should not be seen as an isolated determinant by practitioners, but rather as an additional objective tool to enhance often subjective decisions (McAuley, Baker, et al., 2021).

This study has some limitations that should be acknowledged when interpreting its findings. First, the sample size was small, so the study may have been underpowered and

unable to detect more significant associations, increasing type 2 error. However, highly skilled athletic populations are generally small by nature and are notoriously difficult to gain access to. Indeed, this study's sample (N = 73) is larger than the median sample size (N = 60) reported in a recent review of eighty genetic association studies in football (see McAuley, Hughes, et al., 2021a). Building this research base with studies using transparent methodologies is important so they can contribute to research synthesis approaches in the future to draw more valid and reliable conclusions (McAuley, Baker, et al., 2022). Second, this study did not make adjustments for multiple comparisons, which may have increased type 1 errors. However, due to the exploratory nature of this study, in regards to the novel experimentation methods employed and unique cohort, reducing type 2 errors was considered a priority. This is recommended in exploratory research (see Althouse, 2016), as a main aim is to ensure an important discovery is not missed in the first instance, which can be validated in subsequent dedicated replication studies. Third, the weighting of SNPs and the direction of the allelic associations in the polygenic models were data driven due to the unique population and lack of prior literature using high-powered research designs. As such, inaccurate weightings and opposite scores could have been assigned to specific SNPs and alleles, which may decrease external validity. Fourth, participants were only genotyped for SNPs and epistatic interactions were not considered. There are many other types of genetic polymorphisms associated with psychological phenotypes (e.g., insertions-deletions and copy number variants) and the interactions between genetic variants may alter associations, thus changing the accuracy of polygenic models. Therefore, the addition of more SNPs and other polymorphisms, whilst also considering their interactions, may increase the polygenic models' accuracy. Lastly, due to the

age of the sample the psychological traits assessed may change over time or be different in adults and more validated questionnaires in children may also alter these reported measures

7.5 Conclusion

This study has presented novel evidence regarding the association of inter-individual genetic variation with the mental toughness and personality profiles of youth football players. These findings suggest that in isolation the *DRD3* (rs167771) SNP G allele is associated with lower levels of agreeableness. In addition, the collective influence of all SNPs included in this study were shown to be associated with mental toughness and all personality dimensions except openness. As such, the findings from this exploratory study suggest that several psychological characteristics of academy footballers may be partly determined by genetic factors. Therefore, we suggest that future studies incorporate SNPs associated with psychological variables when conducting research into genetic associations with behavioural traits and/or athletic prowess. Successful independent replication in large homogenous cohorts may allow practitioners in the future to implement more individualised psychological intervention programmes during athlete development, which may help manage athlete welfare and facilitate improved long-term performance.

8 Genetic associations with acceleration, change of direction, jump height, and speed in English academy football players

McAuley, A. B. T., Hughes, D. C., Tsaprouni, L. G., Varley, I., Suraci, B., Baker, J., Herbert, A. J., & Kelly, A. L. (in press). Genetic associations with acceleration, change of direction, jump height, and speed in English academy football players. *The Journal of Strength and Conditioning Research*.

Abstract

High-intensity movements and explosive actions are commonly assessed during athlete development in football (soccer). While many environmental factors underpin these power-orientated traits, research suggests there is also a sizeable genetic component. Therefore, this study examined the association of twenty-two single nucleotide polymorphisms (SNPs) with acceleration, change of direction, jump height, and speed in academy football players. One hundred and forty-nine male under-12 to under-23 football players from four English academies were examined. Participants performed 5 m, 10 m, 20 m, and 30 m sprints, countermovement jumps (CMJs), and the 5-0-5 agility test. Simple linear regression was used to analyse individual SNP associations, whereas both unweighted and weighted total genotype scores (TGSs; TWGSs) were computed to measure the combined influence of all SNPs. In isolation, the *GALNT13* (rs10196189) G allele and *IL6* (rs1800795) G/G genotype were associated with faster (~4%) 5 m, 10 m, and 20 m sprints and higher (~16%) CMJs, respectively. Furthermore, the TGS and TWGS were significantly correlated with all performance assessments, explaining between 6 and 33% of the variance. This study

demonstrates that some genetic variants are associated with power-orientated phenotypes in youth football players and may add value towards a future polygenic profile of physical performance.

8.1 Introduction

Football (soccer) is an intermittent sport that is physiologically characterised by a variety of explosive movements and actions (Bangsbo et al., 2006; Dellal et al., 2011; Mallo et al., 2015). A typical match comprises ~1200 acyclical bursts of activity (e.g., change of direction [COD], jumps, and sprints), with high-intensity running occurring every ~70 seconds (s) (Bradley et al., 2009; Dodd & Newans, 2018). Whilst ~70% of a match requires players to perform low-intensity activities, they travel 10-12 kilometres (km) with an average work intensity of 80% - 90% of maximal heart rate (HR_{max}) and 70% - 80% of maximal oxygen uptake (VO_{2max}), which is close to the anaerobic threshold (Bangsbo et al., 2014; Stølen et al., 2005). Research suggests the frequency of explosive activities performed during a game has rapidly increased. For instance, Barnes et al. (2014) reported that in the English Premier League from the 2006/07 to the 2012/13 season, high-intensity running distance, sprint distance, high-intensity actions, and the number of sprints all increased by ~30%, ~35%, ~50%, and ~85%, respectively. Continued increases in high-intensity running distances and number of sprints have been observed more recently in the Chinese Super League from the 2012/13 to the 2017/18 season (Zhou et al., 2020) and in the Spanish La Liga from the 2015/16 to the 2018/19 season (Pons et al., 2021).

The association of explosive activities with success in football may explain their rise in incidence and quantity. For instance, explosive activities constitute the more crucial moments in a match and directly contribute to winning/losing possession and scoring/conceding goals (Little & Williams, 2005; Oliver et al., 2009). Moreover, higher-intensity actions can distinguish teams of different competitive standards of play and are associated with a team's final placing in their respective league (Di Salvo et al., 2009; Kalapotharakos et al., 2006; Mohr et al., 2003). Indeed, more successful players commonly display greater sprint times, high-

intensity running distances, repeated sprint ability, jump heights, and one-repetition maximum back squat values (Bradley et al., 2009; Stølen et al., 2005). This has been further substantiated by multiple reviews revealing that players of higher competitive levels achieve superior scores on tests which measure acceleration, agility, anaerobic power, COD, speed, and strength (Dodd & Newans, 2018; Haugen et al., 2014; Sarmento et al., 2018; Slimani & Nikolaidis, 2019).

Similar findings have been reported within youth football populations (e.g., Bennett et al., 2019; Gil et al., 2007, 2014; Höner et al., 2021; Kelly et al., 2021; Mirkov et al., 2010; Patel et al., 2020) and have been synthesised in recent reviews (see Kelly & Williams 2020; Vieira et al., 2019). However, from a talent identification perspective, the prognostic value of physiological characteristics possessed at younger ages on performance at adulthood may be of the greatest interest (Murr et al., 2018). For instance, in a longitudinal investigation in academy football players, Saward et al. (2020) reported that future professionals outperformed their non-professional counterparts in countermovement jumps (CMJ) and COD from the age of 12 years, and 20 m sprint throughout their entire development (aged 8-19 years) in England. Several other studies have reported significant prognostic associations with acceleration, speed, COD, and CMJ in different populations and/or specific age groups during development (e.g., Deprez et al., 2015; Emmonds et al., 2016; Forsman et al., 2016; Gonaus & Muller, 2012; Höner et al., 2017; Le Gall et al., 2010; Leyhr et al., 2018).

Due to the contribution of explosive actions to success in football and their potential as significant predictors during athlete development, factors that influence physiological performance are of considerable interest to researchers and practitioners alike (Murr et al., 2018). Several variables affecting physiological performance in youth football have been identified (e.g., stature, body mass, age, maturation, relative age effects, playing position,

flexibility, and training stimuli) (Abarghoueinejad et al., 2021; Dodd & Newans, 2018; Kelly & Williams, 2020; Leyhr et al., 2021; Radnor et al., 2021; Sarmento et al., 2018). However, a largely under-researched component of physiological performance in football is the influence of genetics (McAuley, Hughes, et al., 2021a; McAuley, Hughes, et al., 2022). All observable human traits are affected by the combination of many genetic variants to some extent (Polderman et al., 2015). More specifically, notable heritability estimates have been reported for several phenotypes associated with explosive actions in football (e.g., height = 80%, skeletal muscle mass = 80%, mesomorphy = 80%, explosive anaerobic power = 70%, body mass = 60%, leg strength = 60%) (Calvo et al., 2002; Livshits et al., 2016; Peeters et al., 2007; Silventoinen et al., 2008; Visscher et al., 2006). Indeed, a meta-analysis in this area produced an overall weighted heritability estimate of 52% for 58 strength and power measurements (Zempo et al., 2017), which further highlights the importance of better understanding genetic associations in youth football players.

The evolution and advancement of molecular biology techniques have enabled the analysis of specific genetic markers to account for a proportion of these purported estimates. Indeed, several genetic markers have now been identified which may explain some of the inter-individual variation observed in athletic performance traits (see Ahmetov et al., 2021 for a review). However, the majority of genetic research in athletic performance has focused on individual sports, whilst the studies that have investigated associations in football have predominately comprised case-control designs with senior cohorts (McAuley, Hughes, et al., 2021a, 2021b). These studies have limited practical application in athlete development contexts as they do not reveal which genetic variants are associated with specific performance phenotypes (McAuley, Baker, et al., 2021). Moreover, due to the paucity of cross-sectional and

longitudinal study designs, and the lack of genetic research in youth populations, the effect of genetic variants on physiological traits during development is unclear (Murtagh et al., 2020). As such, the purpose of this study was to investigate the association of twenty-two relevant single nucleotide polymorphisms (SNPs) with acceleration, COD, jump height, and speed in youth football players.

8.2 Methods

8.2.1 Participants

One hundred and forty-nine male under-12 to under-23 (aged 15.72 ± 2.64 years) football players from two Category 1 and Category 3 English academies participated. Informed assent from all players, consent from parents/guardians, and gatekeeper consent from each academy was collected prior to the commencement of the study. All experimental procedures were conducted in accordance with the guidelines in the Declaration of Helsinki and ethical approval was granted by Birmingham City University via the Health, Education, and Life Sciences Academic Ethics Committee. This study was conducted in accordance with the recommendations for reporting the results of genetic association studies defined by the STrengthening the REporting of Genetic Association studies (STREGA) Statement (Little et al., 2009).

8.2.2 Performance tests

The 5 m and 10 m sprint, 5-0-5 agility, CMJ, and 20 m and 30 m sprint tests were used to assess acceleration, COD, jump height, and speed, respectively. All tests have been previously utilised in youth football research and provide valid and reliable assessments (Kelly et al. 2021; Murtagh et al., 2020). All participants were familiarised with the testing procedures before commencement.

For the 5 m, 10 m, 20 m and 30 m sprint tests, light gates (Brower TC Timing System, Draper, Utah, USA) were set up as per the manufacturer's instructions and placed 5 m apart. Participants started approximately 0.5 m behind the first light gate, began the sprint at their own convenience, and continued sprinting until passing the final set of light gates. Three trials were completed, with a 3-minute rest between trials, and the quickest results were used for analysis.

For the 5-0-5 agility test, a light gate (Brower TC Timing System, Draper, Utah, USA) was set up as per the manufacturer's instructions. Participants started approximately 0.5 m behind the light gate and began the test at their own convenience. Participants sprinted towards a line 5 m ahead of the light gate, pivoted 180 degrees, and sprinted past the light gate at the start position. Two trials were completed, turning in different directions in each attempt, with a 3-minute rest between both trials. The quickest result was used for analysis.

For the CMJ test, a jump mat (Just Jump system, Probotics Inc. 8602 Esslinger CT, Huntsville, Alabama, USA) was used. Participants were informed to not swing their arms during the jump and maintain hands on hips. The importance of using a countermovement and the need to take-off and land with straight legs was communicated and demonstrated to each participant. Three trials were completed, with a 3-minute rest between trials, and the greatest jump height was used for analysis.

8.2.3 Genetic procedures

8.2.3.1 Genotyping

Saliva was collected from players via sterile, self-administered buccal swabs, following a minimum of 30 minutes since food or drink ingestion. Saliva samples were safely stored at room temperature and within 36 hours were sent to AKESOgen, Inc. (Peachtree Corners, GA,

USA) for DNA extraction. Using Qiagen chemistry, DNA was extracted on an automated Kingfisher FLEX instrument (Thermo Fisher Scientific, Waltham, MA, US). The manufacturers recommended guidelines and procedures were followed throughout. To measure the extracted DNA's quality and quantity, PicoGreen and Nanodrop measurements were taken. Input to the custom testing array occurs at 200 ng in 20 μL. Amplification, fragmentation, and resuspension were performed using Biomek FXP. GeneTitan instrumentation (Thermo Fisher Scientific, Waltham, MA, US) was used to stain and scan the arrays, with hybridisation performed in a Binder oven at 48 degrees for 24 hours, following the Affymetrix Axiom high throughput 2.0 protocol. Data analysis was then performed using a raw CEL file data input into the Affymetrix Axiom Analysis Suite (Affymetrix, Santa Clara, CA, US). Procedures were in accordance with Pickering et al. (2019).

8.2.3.2 Polymorphism selection

To identify potentially associated polymorphisms, empirical research, review articles, book chapters, and the GWAS catalog (https://www.ebi.ac.uk/gwas/) were examined. Priority was given to GWAS results that were replicated in independent cohorts, followed by candidate gene studies with large homogenous sample sizes which produced similar associations in more than one study. The polymorphisms that were finally included were dependent on the coverage of the microarray and quality control procedures. After an extensive search of the literature, the following twenty-two SNPs were selected based on their proposed biological function and relevant associations in previous studies: Angiotensin I converting enzyme (*ACE*; rs4341), Actinin alpha 3 (*ACTN3*; rs1815739), Adrenoceptor beta 2 (*ADRB2*; rs1042714), Angiotensinogen (*AGT*; rs699), Adenosine monophosphate deaminase 1 (*AMPD1*; rs17602729), Brain derived neurotrophic factor (*BDNF*; rs6265), Creatine kinase, M-type

(*CKM*; rs8111989), Copine 5 (*CPNE5*; rs3213537), FTO alpha-ketoglutarate dependent dioxygenase (*FTO*; rs9939609), Polypeptide N-acetylgalactosaminyltransferase 13 (*GALNT13*; rs10196189), Hypoxia inducible factor 1 subunit alpha (*HIF1A*; rs11549465), Hydroxysteroid 17-beta dehydrogenase 14 (*HSD17B14*; rs7247312), Insulin like growth factor 1 (*IGF1*; rs35767), Insulin like growth factor 2 (*IGF2*; rs680), Interleukin 6 (*IL6*; rs1800795), Nitric oxide synthase 3 (*NOS3*; rs2070744), Peroxisome proliferator activated receptor alpha (*PPARA*; rs4253778), Peroxisome proliferator activated receptor gamma (*PPARG*; rs1801282), Solute carrier family 16 member 1 (*SLC16A1*; rs1049434), Superoxide dismutase 2 (*SOD2*; rs4880), Thyrotropin releasing hormone receptor (*TRHR*; rs7832552), and Uncoupling protein 2 (*UCP2*; rs660339) (see **Table 8.1** for more information). These gene names and symbols are in accordance with those officially approved by the Human Gene Nomenclature Committee (HGNC; https://www.genenames.org). Standard genomic quality control (QC) procedures and thresholds were applied when selecting polymorphisms: SNP call rate (>95), sample call rate (>95), fisher's linear discriminant (>3.6), and minor allele frequency (>0.05).

Table 8.1. Physical gene and single nucleotide polymorphism (SNP) information.

Gene	Symbol	Chr	Function	SNP	Consequence	Associations	MAF
Angiotensin I converting enzyme	*ACE*	17q23.3	Catalyses the degradation of the inactive decapeptide angiotensin I, and generates the physiologically active peptide, angiotensin II.	rs4341	Intron variant C > G (Insertion > Deletion)	Serum and tissue activity, blood pressure, percentage of muscle fibre types, strength and muscle volume.	C = 0.43
Actinin alpha 3	*ACTN3*	11q13.2	Primarily expressed in skeletal muscle and functions as a structural component of sarcomeric Z line.	rs1815739	Nonsense variant C > T (Arg > Ter)	Protein deficiency, percentage of muscle fibre types, strength, muscle volume, acceleration, speed, and power.	T = 0.43
Adrenoceptor beta 2	*ADRB2*	5q32	Expressed in the brain as part of the catecholamine system and acts as a receptor for adrenaline, noradrenaline, and dopamine.	rs1042714	Missense variant G > C (Glu > Gln)	Receptor activity and sensitivity, neurodevelopmental disorders, white matter and cognitive functions.	G = 0.41
Angiotensinogen	*AGT*	1q42.2	Expressed in the liver and is cleaved by the enzyme renin in response to lowered blood pressure and produces angiotensin I.	rs699	Missense variant A > G (Met > Thr)	Plasma levels, serum concentration, blood pressure, skeletal muscle hypertrophy and volume.	G = 0.41
Adenosine monophosphate deaminase 1	*AMPD1*	1p13.2	Catalyses the deamination of AMP to IMP in skeletal muscle.	rs17602729	Nonsense variant G > A (Gln > Ter)	Enzyme functionality, AMP metabolism, muscle fatigue, weakness and cramping.	A = 0.12

Brain derived neurotrophic factor	*BDNF*	11p14.1	Encodes for a neurotrophin essential for neuronal development and survival, synaptic plasticity, and cognitive function.	rs6265	Missense variant C > T (Val > Met)	Motor cortex plasticity, motor learning, neuromuscular activation, horizontal and vertical power.	T = 0.20
Creatine kinase, M-type	*CKM*	19q13.32	A cytoplasmic enzyme involved in energy homeostasis and an important serum marker for myocardial infarction.	rs8111989	500B Downstream variant T > C	Skeletal muscle performance, maximum oxygen uptake, trainability, power and endurance athlete status.	C = 0.30
Copine 5	*CPNE5*	6p21.2	Calcium-dependent membrane-binding proteins may regulate molecular events at the interface of the cell membrane and cytoplasm	rs3213537	Intron variant C > T	Percentage of muscle fibre types, acceleration, speed, and power athlete status.	T = 0.14
FTO alpha-ketoglutarate dependent dioxygenase	*FTO*	16q12.2	Expressed in the hypothalamic nuclei that regulate energy balance.	rs9939609	Intron variant T > A	Fat mass, lean mass, muscle power, percentage of muscle fibre types, and power and endurance athlete status.	A = 0.41
Polypeptide N-acetylgalactosaminyltransferase 13	*GALNT13*	2q23.3-q24.1	Expressed in the brain, B cells, kidney, and liver as well as potentially being involved in metabolism and energy pathways.	rs10196189	Intron variant A > G	Power athlete status and speed.	G = 0.14
Hypoxia inducible factor 1 subunit alpha	*HIF1A*	14q23.2	Regulates cellular and systemic homeostatic response to hypoxia by activating transcription of many genes.	rs11549465	Missense variant C > T (Pro > Ser)	Protein stability, transcriptional activity, percentage of muscle fibre types, and power and endurance athlete status.	T = 0.10

Hydroxysteroid 17-beta dehydrogenase 14	*HSD17B14*	19q13.33	Involved in metabolism of steroids at the C17 position and also of other substrates, such as fatty acids, prostaglandins, and xenobiotics.	rs7247312	Intron variant A > G	Power athlete status, acceleration, and speed.	G = 0.10
Insulin like growth factor 1	*IGF1*	12q23.2	Protein similar to insulin in function and structure involved in mediating growth and development.	rs35767	Missense variant G > A (Gly > Val)	Circulating IGF1 levels, body composition, power athlete status, and strength.	A = 0.16
Insulin like growth factor 2	*IGF2*	11p15.5	A member of the insulin family of polypeptide growth factors, which are involved in development and growth.	rs680	3 Prime UTR variant C > T	Acceleration, speed, and power athlete status.	T = 0.32
Interleukin 6	*IL6*	7p15.3	A pleiotropic cytokine involved in glucose homeostasis, hypertrophic muscle growth, immune function, and muscle damage repair.	rs1800795	Intron variant G > C	Acceleration, speed, and power and strength athlete status.	C = 0.42
Nitric oxide synthase 3	*NOS3*	7q36.1	A biologic mediator in several processes, potentially stimulating muscle hypertrophy through NO-mediated vasodilatation.	rs2070744	Intron variant C > T	Promoter activity, endothelial nitric oxide synthesis, muscle hypertrophy, acceleration, and power athlete status.	C = 0.44
Peroxisome proliferator activated receptor alpha	*PPARA*	22q13.31	Present in skeletal muscle and promotes uptake, utilization, and catabolism of fatty acids by upregulation of other genes.	rs4253778	Intron variant G > C	Percentage of muscle fibre types, strength, power, muscle mass, and power and endurance athlete status.	C = 0.19

Peroxisome proliferator activated receptor gamma	*PPARG*	3p25.2	A central transcriptional regulator of adipogenic and lipogenic programs, insulin sensitivity and glucose homeostasis	rs1801282	Missense variant C > G (Pro > Ala)	Receptor activity, insulin sensitivity, skeletal muscle glucose uptake, cross-sectional area of muscle fibres.	G = 0.12
Solute carrier family 16 member 1	*SLC16A1*	1p13.2	A monocarboxylate transporter that catalyses the movement of monocarboxylates, such as lactate and pyruvate.	rs1049434	Missense variant T > A (Asp > Glu)	Lactate transport rate, lactate accumulation, maximum oxygen uptake, power and endurance athlete status.	A = 0.44
Superoxide dismutase 2	*SOD2*	6q25.3	Catalyses the dismutation of superoxide radicals in mitochondria by converting anion superoxide into hydrogen peroxide and oxygen	rs4880	Missense variant A > G (Val > Ala)	MnSOD efficiency against oxidative stress, MnSOD protein formation, CKM activity, power athlete status.	G = 0.47
Thyrotropin releasing hormone receptor	*TRHR*	8q23.1	Stimulates the release of thyrotropin and prolactin, having a role in the regulation of metabolic and hormonal functions.	rs7832552	Intron variant C > T	Lean body mass, strength, and power athlete status.	T = 0.27
Uncoupling protein 2	*UCP2*	11q13.4	Involved in energy expenditure that facilitates the transfer of anions and protons in the inner and outer mitochondrial membrane.	rs660339	Missense variant G > A (Ala > Val)	Energy expenditure, body mass, aerobic and anaerobic capacities, endurance and power athlete status.	A = 0.40

Note. Chr = chromosome location; MAF = minor allele frequency (according to European population; 1000 Genomes Project Consortium, 2015).

8.2.3.3 Total genotype score

Unweighted and weighted total genotype scores (TGS; TWGS) were calculated to assess the combined influence of the included SNPs on each physical performance test. Both TGSs and TWGSs have demonstrated sufficient discriminatory power in previous sport genomic research (Charlier et al., 2017; Massidda et al., 2014; Varillas Delgado et al., 2020). To generate both the TGS and TWGS, each genotype of a respective SNP initially received a score between 0-2 based on the observed genotype associations with a dependent variable. Genotypes of dominant (AA vs. Aa-aa) and recessive (AA-Aa vs. aa) models were assigned a score of two (i.e., associated genotype[s]) or zero (i.e., alternate genotype[s]), whereas genotypes of co-dominant models (AA vs. Aa vs. aa) were assigned three scores (i.e., homozygous-associated genotypes received a score of two, the heterozygote received a score of one, and the alternate homozygous genotype received a score of zero) (Guilherme et al., 2020).

For the TGS, the original procedure of Williams and Folland (2008) was followed. Genotype scores (GS) were summed and transformed into a 0-100 scale by dividing the total score by the maximum possible score and multiplying by 100.

$$TGS = (combined\text{-}GS / maximum\text{-}GS) * 100$$

For the TWGS, a similar procedure to Varillas Delgado et al. (2020) was used. Each GS was multiplied by the standardised beta coefficients (β) of each SNP following multiple regression with each dependent variable to create weighted genotype scores (WGS). The WGSs were then summed and transformed into a 0-100 scale by dividing the total score by the maximum possible score and multiplying by 100.

$$TWGS = (combined\text{-}WGS / maximum\text{-}WGS) * 100$$

8.2.4 Data analysis

Each SNP was tested for adherence with Hardy-Weinberg equilibrium (HWE) using an exact test via SNPStats (Solé et al., 2006). Linkage disequilibrium (LD) was analysed using LDlink (Machiela & Chanock, 2015) and data from the 1000 Genomes Project European ancestry population (1000 Genomes Project Consortium, 2015). All other data were analysed using Jamovi version 1.8.1 and IBM SPSS version 25. Normality was assessed with the Kolmogorov–Smirnov test and homoscedasticity was assessed using Levene's test. Akaike information criterion (AIC) was used to select which genetic model (i.e., co-dominant, dominant, recessive) best fits the data and would be subjected to hypothesis testing. However, if MAF ≤ 0.25 a dominant model was utilised to retain statistical power (Murtagh et al., 2020). Simple linear regression with age added as a covariate was performed to assess the association of genotype models with each performance test. To control for multiple testing, a Benjamini–Hochberg false discovery rate (FDR) of 0.05 was applied to all (N = 132) genotype model comparisons (Benjamini & Hochberg, 1995). Multiple regression was used to calculate the standardised beta coefficients (β) of each SNP for the TWGS models. Simple linear regression was then performed to assess the association of each TGS and TWGS with each dependent variable. Pearson's correlation coefficient (r) with threshold values of ≤0.1 (trivial), >0.1-0.3 (small), >0.3–0.5 (moderate), >0.5–0.7 (large), >0.7–0.9 (very large), and >0.9–1.0 (almost perfect) were used to measure correlation (Hopkins, 2009). The coefficient of determination (R^2) was computed to determine the variance explained by each TGS and TWGS. Statistical significance was set at $p < .05$.

8.3 Results

Genotype and allele distributions of all SNPs were in HWE, except for *GALNT13* ($p < .001$) and *UCP2* ($p = .030$), and all SNPs were in linkage equilibrium. Assumptions of normality and homoscedasticity were not violated. Descriptive statistics and genotype frequencies are displayed in **Table 8.2**.

Table 8.2. Descriptive statistics of physical assessments.

Gene (SNP)	Genotype = *n* (%)	5 m (s)	10 m (s)	20 m (s)	30 m (s)	5-0-5 (s)	CMJ (cm)	MAF	HWE
Overall		1.04 ± 0.08	1.79 ± 0.17	3.11 ± 0.32	4.46 ± 0.35	2.39 ± 0.13	39.82 ± 7.63	N/A	N/A
ACE (rs4341)	G/G = 39 (26)	1.06 ± 0.10	1.85 ± 0.14	3.19 ± 0.34	4.58 ± 0.38	2.39 ± 0.13	38.95 ± 8.83	0.49	1
	G/C = 75 (50)	1.02 ± 0.08	1.76 ± 0.18	3.07 ± 0.27	4.40 ± 0.32	2.39 ± 0.13	40.37 ± 6.33		
	C/C = 35 (23)	1.04 ± 0.08	1.81 ± 0.13	3.10 ± 0.36	4.49 ± 0.35	2.40 ± 0.12	39.68 ± 8.69		
ACTN3 (rs1815739)	C/C = 55 (37)	1.02 ± 0.08	1.78 ± 0.17	3.10 ± 0.25	4.40 ± 0.65	2.39 ± 0.12	39.31 ± 7.64	0.38	0.49
	C/T = 75 (50)	1.04 ± 0.09	1.80 ± 0.17	3.14 ± 0.30	4.50 ± 0.34	2.39 ± 0.13	40.24 ± 7.90		
	T/T = 19 (13)	1.05 ± 0.09	1.82 ± 0.13	2.98 ± 0.49	4.54 ± 0.36	2.39 ± 0.10	39.55 ± 6.99		
ADBR2 (rs1042714)	C/C = 43 (29)	1.03 ± 0.09	1.81 ± 0.14	3.15 ± 0.25	4.45 ± 0.36	2.39 ± 0.12	38.5 ± 7.63	0.48	0.41
	C/G = 69 (46)	1.05 ± 0.09	1.79 ± 0.20	3.12 ± 0.35	4.52 ± 0.35	2.40 ± 0.13	40.38 ± 8.02		
	G/G = 37 (25)	1.02 ± 0.08	1.79 ± 0.13	3.03 ± 0.33	4.38 ± 0.32	2.38 ± 0.32	40.01 ± 7.13		
AGT (rs699)	A/A = 47 (32)	1.04 ± 0.08	1.78 ± 0.20	3.07 ± 0.38	4.52 ± 0.36	2.38 ± 0.11	39.84 ± 6.57	0.45	0.41

	A/G = 69 (46)	1.03 ± 0.09	1.79 ± 0.16	3.12 ± 0.31	4.43 ± 0.35	2.38 ± 0.12	38.60 ± 8.41		
	G/G = 33 (22)	1.02 ± 0.07	1.82 ± 0.12	3.13 ± 0.22	4.47 ± 0.32	2.43 ± 0.15	41.95 ± 7.43		
	G/G = 119 (80)	1.04 ± 0.08	1.80 ± 0.16	3.11 ± 0.34	4.47 ± 0.35	2.40 ± 0.13	40.57 ± 7.72		
AMPD1 (rs17602729)	G/A = 29 (19)	1.02 ± 0.09	1.76 ± 0.17	3.09 ± 0.23	4.42 ± 0.32	2.35 ± 0.11	37.11 ± 6.18	0.10	1
	A/A = 1 (1)	1.06 ± N/A	1.89 ± N/A	3.39 ± N/A	4.87 ± N/A	2.40 ± N/A	27.50 ± N/A		
	C/C = 99 (66)	1.04 ± 0.09	1.80 ± 0.17	3.12 ± 0.29	4.47 ± 0.37	2.40 ± 0.13	39.62 ± 6.83		
BDNF (rs6265)	C/T = 45 (30)	1.02 ± 0.07	1.79 ± 0.16	3.08 ± 0.38	4.46 ± 0.30	2.38 ± 0.13	40.66 ± 8.64	0.18	1
	T/T = 5 (3)	1.00 ± 0.08	1.75 ± 0.11	3.09 ± 0.23	4.34 ± 0.32	2.37 ± 0.04	35.91 ± 10.12		
	T/T = 80 (54)	1.02 ± 0.08	1.78 ± 0.14	3.06 ± 0.32	4.43 ± 0.32	2.40 ± 0.12	40.20 ± 7.83		
CKM (rs8111989)	T/C = 55 (37)	1.05 ± 0.08	1.81 ± 0.19	3.15 ± 0.31	4.48 ± 0.36	2.37 ± 0.13	39.37 ± 7.07	0.28	0.31
	C/C = 14 (9)	1.04 ± 0.11	1.80 ± 0.18	3.16 ± 0.32	4.62 ± 0.44	2.41 ± 0.12	39.16 ± 9.26		
	C/C = 103 (69)	1.04 ± 0.09	1.79 ± 0.17	3.13 ± 0.25	4.47 ± 0.35	2.39 ± 0.13	40.23 ± 8.04		
CPNE5 (rs3213537)	C/T = 43 (29)	1.03 ± 0.08	1.79 ± 0.17	3.08 ± 0.39	4.46 ± 0.36	2.39 ± 0.13	38.40 ± 6.35	0.16	0.77
	T/T = 3 (2)	1.16 ± N/A	1.89 ± 0.03	2.07 ± N/A	4.46 ± 0.14	2.48 ± 0.13	42.39 ± 9.20		
	T/T = 45 (30)	1.05 ± 0.09	1.81 ± 0.14	3.16 ± 0.24	4.46 ± 0.35	2.37 ± 0.12	39.35 ± 7.70		
FTO (rs9939609)	T/A = 78 (52)	1.03 ± 0.08	1.77 ± 0.19	3.09 ± 0.33	4.47 ± 0.35	2.40 ± 0.13	39.47 ± 7.65	0.44	0.51
	A/A = 26 (17)	1.04 ± 0.08	1.82 ± 0.13	3.08 ± 0.40	4.47 ± 0.36	2.41 ± 0.12	41.71 ± 7.69		

	A/A = 99 (66)	1.05 ± 0.08	1.81 ± 0.18	3.16 ± 0.33	4.51 ± 0.36	2.39 ± 0.13	39.37 ± 6.85		
GALNT13 (rs10196189)	A/G = 34 (23)	1.03 ± 0.09	1.80 ± 0.12	3.06 ± 0.32	4.50 ± 0.28	2.41 ± 0.12	40.95 ± 9.36	0.22	<0.001
	G/G = 16 (11)	0.95 ± 0.04	1.67 ± 0.07	2.94 ± 0.12	4.13 ± 0.18	N/A ± N/A	N/A ± N/A		
	C/C = 116 (78)	1.04 ± 0.09	1.81 ± 0.16	3.14 ± 0.28	4.50 ± 0.35	2.39 ± 0.12	39.36 ± 7.31		
HIF1A (rs11549465)	C/T = 32 (21)	1.02 ± 0.07	1.75 ± 0.16	2.99 ± 0.42	4.36 ± 0.31	2.39 ± 0.13	41.39 ± 8.93	0.11	0.69
	T/T = 1 (1)	N/A ± N/A	1.89 ± N/A	N/A ± N/A	4.45 ± N/A	2.54 ± N/A	43.10 ± N/A		
	A/A = 120 (81)	1.03 ± 0.09	1.78 ± 0.17	3.08 ± 0.33	4.45 ± 0.34	2.40 ± 0.13	39.84 ± 7.12		
HSD17B14 (rs7247312)	A/G = 26 (17)	1.05 ± 0.08	1.83 ± 0.13	3.18 ± 0.24	4.48 ± 0.37	2.38 ± 0.11	40.95 ± 9.84	0.11	0.38
	G/G = 3 (2)	1.09 ± 0.08	1.93 ± 0.14	3.39 ± 0.23	4.80 ± 0.30	2.35 ± 0.11	31.45 ± 6.29		
	G/G = 98 (66)	1.05 ± 0.08	1.82 ± 0.16	3.14 ± 0.32	4.51 ± 0.36	2.39 ± 0.13	39.58 ± 7.13		
IGF1 (rs35767)	G/A = 47 (32)	1.01 ± 0.08	1.76 ± 0.14	3.04 ± 0.31	4.37 ± 0.31	2.41 ± 0.13	40.45 ± 8.75	0.18	0.79
	A/A = 4 (3)	1.02 ± 0.11	1.56 ± 0.35	3.04 ± 0.18	4.35 ± 0.17	2.30 ± 0.10	37.30 ± 6.22		
	C/C = 77 (52)	1.05 ± 0.08	1.81 ± 0.18	3.17 ± 0.31	4.51 ± 0.35	2.40 ± 0.13	40.21 ± 7.94		
IGF2 (rs680)	C/T = 65 (44)	1.02 ± 0.09	1.77 ± 0.13	3.05 ± 0.30	4.41 ± 0.35	2.38 ± 0.12	40.00 ± 7.23	0.27	0.21
	T/T = 7 (5)	1.09 ± 0.08	1.79 ± 0.29	2.96 ± 0.45	4.47 ± 0.27	2.39 ± 0.11	31.54 ± 1.98		
IL6 (rs1800795)	G/G = 52 (35)	1.03 ± 0.08	1.80 ± 0.17	3.10 ± 0.34	4.45 ± 0.35	2.40 ± 0.13	44.64 ± 7.35	0.41	0.87
	G/C = 71 (48)	1.03 ± 0.09	1.78 ± 0.18	3.09 ± 0.28	4.47 ± 0.34	2.40 ± 0.13	38.52 ± 6.11		

	C/C = 26 (17)	1.04 ± 0.07	1.82 ± 0.13	3.15 ± 0.36	4.49 ± 0.37	2.36 ± 0.09	34.36 ± 7.68		
	T/T = 58 (39)	1.04 ± 0.09	1.82 ± 0.13	3.12 ± 0.30	4.47 ± 0.36	2.39 ± 0.13	38.93 ± 7.09		
NOS3 (rs2070744)	T/C = 73 (49)	1.03 ± 0.08	1.78 ± 0.19	3.11 ± 0.30	4.47 ± 0.34	2.39 ± 0.12	40.47 ± 7.33	0.37	0.60
	C/C = 18 (12)	1.04 ± 0.10	1.78 ± 0.13	3.06 ± 0.43	4.42 ± 0.33	2.39 ± 0.12	39.81 ± 10.46		
	G/G = 96 (64)	1.03 ± 0.08	1.79 ± 0.14	3.08 ± 0.30	4.43 ± 0.32	2.38 ± 0.12	40.17 ± 7.84		
PPARA (rs4253778)	G/C = 45 (30)	1.05 ± 0.09	1.81 ± 0.22	3.18 ± 0.36	4.57 ± 0.39	2.41 ± 0.14	38.37 ± 7.15	0.20	0.45
	C/C = 8 (5)	0.99 ± 0.08	1.76 ± 0.10	3.02 ± 0.14	4.27 ± 0.13	2.44 ± 0.18	46.30 ± 4.80		
	C/C = 123 (83)	1.04 ± 0.09	1.80 ± 0.17	3.11 ± 0.34	4.48 ± 0.36	2.40 ± 0.12	39.40 ± 8.13		
PPARG (rs1801282)	C/G = 24 (16)	1.03 ± 0.05	1.76 ± 0.17	3.11 ± 0.17	4.39 ± 0.26	2.36 ± 0.14	41.02 ± 5.66	0.09	0.62
	G/G = 2 (1)	1.06 ± 0.05	1.78 ± 0.06	3.06 ± 0.10	4.24 ± 0.12	2.30 ± 0.01	45.50 ± N/A		
	T/T = 50 (34)	1.03 ± 0.09	1.79 ± 0.17	3.07 ± 0.34	4.44 ± 0.33	2.40 ± 0.14	41.35 ± 8.73		
SLC16A1 (rs1049434)	T/A = 74 (50)	1.03 ± 0.09	1.80 ± 0.16	3.13 ± 0.30	4.48 ± 0.38	2.39 ± 0.12	39.54 ± 6.67	0.42	0.87
	A/A = 50 (34)	1.05 ± 0.08	1.78 ± 0.19	3.09 ± 0.32	4.47 ± 0.28	2.38 ± 0.10	37.35 ± 7.06		
	A/A = 40 (27)	1.06 ± 0.09	1.83 ± 0.13	3.12 ± 0.32	4.50 ± 0.34	2.39 ± 0.13	38.62 ± 5.95		
SOD2 (rs4880)	A/G = 73 (49)	1.03 ± 0.08	1.79 ± 0.16	3.10 ± 0.36	4.46 ± 0.35	2.40 ± 0.13	42.23 ± 8.42	0.49	0.87
	G/G = 36 (24)	1.01 ± 0.07	1.76 ± 0.21	3.11 ± 0.22	4.44 ± 0.34	2.39 ± 0.12	36.20 ± 5.88		
TRHR (rs7832552)	C/C = 78 (52)	1.03 ± 0.08	1.80 ± 0.13	3.10 ± 0.25	4.44 ± 0.35	2.40 ± 0.13	40.34 ± 8.15	0.30	0.12

	C/T = 54 (36)	1.05 ± 0.09	1.80 ± 0.20	3.10 ± 0.42	4.51 ± 0.34	2.39 ± 0.13	38.84 ± 7.22		
	T/T 17 (11)	1.04 ± 0.10	1.76 ± 0.20	3.15 ± 0.25	4.45 ± 0.35	2.37 ± 0.09	41.64 ± 6.54		
	G/G = 37 (25)	1.04 ± 0.09	1.83 ± 0.15	3.13 ± 0.39	4.47 ± 0.42	2.40 ± 0.13	42.07 ± 7.00		
UCP2 (rs660339)	G/A = 88 (59)	1.03 ± 0.08	1.77 ± 0.18	3.10 ± 0.27	4.45 ± 0.32	2.40 ± 0.13	39.19 ± 7.52	0.46	0.03
	A/A = 24 (16)	1.04 ± 0.09	1.82 ± 0.13	3.09 ± 0.37	4.51 ± 0.32	2.37 ± 0.09	38.98 ± 8.89		

Note. Data presented in mean ± standard deviation. CMJ = countermovement jump; MAF = minor allele frequency; HWE = Hardy-Weinberg equilibrium.

8.3.1 Individual SNPs

8.3.1.1 5 m and 10 m sprint

There were significant associations between *ACTN3* ($F_{(1, 120)} = 5.95$, $p = .016$), *CKM* ($F_{(1, 120)} = 4.51$, $p = .036$), *GALNT13* ($F_{(1, 120)} = 14.22$, $p < .001$), *SOD2* ($F_{(1, 120)} = 4.71$, $p = .032$) and 5m sprint times (see **Table 8.3**). Significant associations were also found between *ACE* ($F_{(1, 143)} = 4.32$, $p = .039$), *CKM* ($F_{(1, 143)} = 5.88$, $p = .017$), *GALNT13* ($F_{(1, 143)} = 14.67$, $p < .001$), *HIF1A* ($F_{(1, 143)} = 5.41$, $p = .021$), *IGF1* ($F_{(1, 143)} = 5.42$, $p = .021$) and 10 m sprint times. However, following FDR correction, only *GALNT13* remained significant, with G allele carriers 4.9% ($B = -0.05$) and 3.3% ($B = -0.06$) faster than A/A homozygotes in 5 m and 10 m sprints, respectively.

8.3.1.2 20 m and 30 m sprint

There was also a significant association between *GALNT13* and 20 m sprint times ($F_{(1, 117)} = 18.66$, $p < .001$), with G allele carriers 4.2% faster than A/A homozygotes ($B = -0.13$). Other significant associations were observed with *ACE* ($F_{(1, 117)} = 4.24$, $p = .042$), *CKM* ($F_{(1, 117)} =$

7.24, p = .008), *IGF2* ($F_{(1, 117)}$ = 5.89, p = .017) and 20 m sprint times, as well as *ACE* ($F_{(1, 130)}$ = 5.89, p = .017), *GALNT13* ($F_{(1, 130)}$ = 8.55, p = .004), *IGF1* ($F_{(1, 130)}$ = 5.74, p = .018) and 30 m sprint times, but these did not remain significant following FDR correction.

8.3.1.3 5-0-5 agility

There were no significant associations between any single SNP and 5-0-5 agility performance, even before FDR correction.

8.3.1.4 Countermovement jump

There were significant associations between *AMPD1* ($F_{(1, 84)}$ = 6.01, p = .016), *IL6* ($F_{(1, 84)}$ = 17.72, p < .001) and CMJ performance. However, following FDR correction, only *IL6* remained significant, with G/G homozygotes jumping 15.7% higher than C allele carriers (B = 6.42).

8.3.2 TGS

The TGS was significantly associated with 5 m ($F_{(1, 121)}$ = 20.34, p < .001) 10 m ($F_{(1, 144)}$ = 26.06, p < .001), 20 m ($F_{(1, 118)}$ = 17.85, p < .001), and 30 m ($F_{(1, 131)}$ = 15.23, p < .001) sprints, as well as 5-0-5 agility ($F_{(1, 106)}$ = 6.84, p = .010) and CMJ performance ($F_{(1, 85)}$ = 18.38, p < .001). More specifically, moderate negative correlations were found with 5 m (r = -.38; R^2 = .14), 10 m (r = -.39; R^2 = .15), 20 m (r = -.36; R^2 = .13), and 30 m (r = -.32; R^2 = .10) sprint times, whilst a small negative correlation with 5-0-5 agility (r = -.25; R^2 = .06) and a moderate positive correlation with CMJ performance (r = .42; R^2 = .18) were shown (see **Figure 8.1**, **Figure 8.2**, and **Figure 8.3**).

8.3.3 TWGS

There were significant associations between the TWGS and 5 m ($F_{(1, 121)}$ = 29.93, p < .001) 10 m ($F_{(1, 144)}$ = 38.47, p < .001), 20 m ($F_{(1, 118)}$ = 36.15, p < .001), and 30 m ($F_{(1, 131)}$ = 28.71, p

< .001) sprints, as well as 5-0-5 agility ($F_{(1, 106)} = 10.79$, $p = .001$) and CMJ performance ($F_{(1, 85)} = 41.66$, $p < .001$). More specifically, moderate negative correlations were found with 5 m ($r = -.45$; $R^2 = .20$), 10 m ($r = -.46$; $R^2 = .21$), 20 m ($r = -.48$; $R^2 = .23$), and 30 m ($r = -.42$; $R^2 = .18$) sprint times, whilst a small negative correlation with 5-0-5 agility ($r = -.30$; $R^2 = .09$) and a large positive correlation with CMJ performance ($r = .57$; $R^2 = .33$) were shown (see **Figure 8.4**, **Figure 8.5**, and **Figure 8.6**).

Table 8.3. Physical simple linear regression analysis.

Gene (SNP)	Model	*B*	*SE B*	β	*t*	*p*
5 m						
ACE (rs4341)	C/C-C/G vs. G/G	-0.02	0.01	-0.28	-1.72	.088
ACTN3 (rs1815739)	C/C vs. C/T-T/T	-0.03	0.01	-0.37	-2.44	**.016***
ADBR2 (rs1042714)	C/C vs. C/G-G/G	-0.01	0.01	-0.17	-1.05	.294
AGT (rs699)	G/G-G/A vs. A/A	-0.02	0.01	-0.27	-1.74	.084
AMPD1 (rs17602729)	G/A-A/A vs. G/G	-0.01	0.01	-0.07	-0.41	.680
BDNF (rs6265)	T/T-T/C vs. C/C	-0.03	0.01	-0.30	-1.98	.050
CKM (rs8111989)	T/T vs. T/C-C/C	-0.03	0.01	-0.30	-2.12	**.036***
CPNE5 (rs3213537)	T/T-T/C vs. C/C	-0.01	0.01	-0.06	-0.38	.702
FTO (rs9939609)	T/A-A/A vs. T/T	-0.02	0.01	-0.18	-1.18	.240
GALNT13 (rs10196189)	G/G-G/A vs. A/A	-0.05	0.01	-0.55	-3.77	**< .001****
HIF1A (rs11549465)	T/T-T/C vs. C/C	-0.03	0.01	-0.32	-1.79	.075
HSD17B14 (rs7247312)	A/A vs. A/G-G/G	0.00	0.02	0.00	0.00	.997
IGF1 (rs35767)	A/A-A/G vs. G/G	-0.02	0.01	-0.28	-1.82	.071

IGF2 (rs680)	T/T-T/C vs. C/C	-0.01	0.01	-0.12	-0.83	.409
IL6 (rs1800795)	G/G vs. G/C-C/C	0.00	0.01	-0.02	-0.15	.885
NOS3 (rs2070744)	C/C-C/T vs. T/T	0.00	0.01	0.00	-0.01	.993
PPARA (rs4253778)	C/C-C/G vs. G/G	0.00	0.01	-0.03	-0.17	.864
PPARG (rs1801282)	G/G-G/C vs. C/C	-0.01	0.02	-0.07	-0.34	.732
SLC16A1 (rs1049434)	T/T vs. T/A-A/A	0.00	0.01	-0.02	-0.13	.897
SOD2 (rs4880)	G/G vs. A/A-A/G	-0.03	0.01	-0.35	-2.17	**.032***
TRHR (rs7832552)	C/C vs. C/T-T/T	-0.02	0.01	-0.21	-1.47	.145
UCP2 (rs660339)	G/G vs. G/A-A/A	0.00	0.01	-0.05	-0.30	.762
10 m						
ACE (rs4341)	C/C-C/G vs. G/G	-0.04	0.02	-0.28	-2.08	**.039***
ACTN3 (rs1815739)	C/C vs. C/T-T/T	-0.03	0.02	-0.21	-1.67	.098
ADBR2 (rs1042714)	C/C vs. C/G-G/G	-0.01	0.02	-0.05	-0.41	.683
AGT (rs699)	G/G-G/A vs. A/A	-0.01	0.02	-0.11	-0.80	.422
AMPD1 (rs17602729)	G/A-A/A vs. G/G	0.00	0.02	-0.04	-0.24	.808
BDNF (rs6265)	T/T-T/C vs. C/C	-0.02	0.02	-0.12	-0.97	.334
CKM (rs8111989)	T/T vs. T/C-C/C	-0.04	0.02	-0.29	-2.42	**.017***
CPNE5 (rs3213537)	T/T-T/C vs. C/C	0.00	0.02	0.00	0.03	.979
FTO (rs9939609)	T/A-A/A vs. T/T	-0.01	0.02	-0.05	-0.38	.706
GALNT13 (rs10196189)	G/G-G/A vs. A/A	-0.06	0.02	-0.47	-3.83	**< .001****
HIF1A (rs11549465)	T/T-T/C vs. C/C	-0.04	0.02	-0.34	-2.33	**.021***
HSD17B14 (rs7247312)	A/A vs. A/G-G/G	-0.01	0.02	-0.10	-0.63	.529

IGF1 (rs35767)	A/A-A/G vs. G/G	-0.04	0.02	-0.29	-2.33	**.021***
IGF2 (rs680)	T/T-T/C vs. C/C	-0.03	0.02	-0.24	-1.96	.052
IL6 (rs1800795)	G/G vs. G/C-C/C	0.01	0.02	0.08	0.60	.552
NOS3 (rs2070744)	C/C-C/T vs. T/T	-0.02	0.02	-0.13	-1.09	.279
PPARA (rs4253778)	C/C-C/G vs. G/G	0.01	0.02	0.10	0.79	.432
PPARG (rs1801282)	G/G-G/C vs. C/C	-0.03	0.02	-0.20	-1.23	.219
SLC16A1 (rs1049434)	T/T vs. T/A-A/A	0.00	0.02	-0.02	-0.17	.866
SOD2 (rs4880)	G/G vs. A/A-A/G	-0.03	0.02	-0.25	-1.76	.080
TRHR (rs7832552)	C/C vs. C/T-T/T	-0.02	0.02	-0.15	-1.24	.216
UCP2 (rs660339)	G/G vs. G/A-A/A	0.01	0.02	0.07	0.49	.622
20 m						
ACE (rs4341)	C/C-C/G vs. G/G	-0.07	0.04	-0.29	-2.06	**.042***
ACTN3 (rs1815739)	C/C vs. C/T-T/T	-0.06	0.03	-0.24	-1.89	.062
ADBR2 (rs1042714)	C/C vs. C/G-G/G	0.00	0.03	-0.01	-0.04	.969
AGT (rs699)	G/G-G/A vs. A/A	-0.02	0.03	-0.08	-0.57	.572
AMPD1 (rs17602729)	G/A-A/A vs. G/G	0.00	0.04	0.02	0.12	.907
BDNF (rs6265)	T/T-T/C vs. C/C	-0.01	0.03	-0.05	-0.37	.715
CKM (rs8111989)	T/T vs. T/C-C/C	-0.08	0.03	-0.32	-2.69	**.008***
CPNE5 (rs3213537)	T/T-T/C vs. C/C	0.01	0.03	0.03	0.21	.838
FTO (rs9939609)	T/A-A/A vs. T/T	-0.03	0.03	-0.11	-0.86	.391
GALNT13 (rs10196189)	G/G-G/A vs. A/A	-0.13	0.03	-0.53	-4.32	**< .001****
HIF1A (rs11549465)	T/T-T/C vs. C/C	-0.07	0.04	-0.28	-1.81	.072

HSD17B14 (rs7247312)	A/A vs. A/G-G/G	-0.02	0.04	-0.06	-0.39	.694
IGF1 (rs35767)	A/A-A/G vs. G/G	-0.06	0.03	-0.24	-1.87	.063
IGF2 (rs680)	T/T-T/C vs. C/C	-0.07	0.03	-0.30	-2.43	**.017***
IL6 (rs1800795)	G/G vs. G/C-C/C	0.00	0.03	0.02	0.13	.899
NOS3 (rs2070744)	C/C-C/T vs. T/T	-0.01	0.03	-0.05	-0.40	.687
PPARA (rs4253778)	C/C-C/G vs. G/G	0.03	0.03	0.12	0.95	.346
PPARG (rs1801282)	G/G-G/C vs. C/C	-0.05	0.04	-0.19	-1.15	.254
SLC16A1 (rs1049434)	T/T vs. T/A-A/A	-0.02	0.03	-0.10	-0.71	.481
SOD2 (rs4880)	G/G vs. A/A-A/G	-0.04	0.04	-0.17	-1.23	.223
TRHR (rs7832552)	C/C vs. C/T-T/T	-0.05	0.03	-0.22	-1.77	.080
UCP2 (rs660339)	G/G vs. G/A-A/A	0.01	0.04	0.06	0.38	.706
30 m						
ACE (rs4341)	C/C-C/G vs. G/G	-0.10	0.04	-0.30	-2.43	**.017***
ACTN3 (rs1815739)	C/C vs. C/T-T/T	-0.06	0.04	-0.16	-1.42	.159
ADBR2 (rs1042714)	C/C vs. C/G-G/G	-0.04	0.04	-0.13	-1.06	.293
AGT (rs699)	G/G-G/A vs. A/A	-0.05	0.04	-0.13	-1.11	.269
AMPD1 (rs17602729)	G/A-A/A vs. G/G	0.00	0.05	-0.01	-0.05	.957
BDNF (rs6265)	T/T-T/C vs. C/C	-0.04	0.04	-0.13	-1.08	.284
CKM (rs8111989)	T/T vs. T/C-C/C	-0.06	0.04	-0.17	-1.56	.121
CPNE5 (rs3213537)	T/T-T/C vs. C/C	0.00	0.04	0.01	0.06	.953
FTO (rs9939609)	T/A-A/A vs. T/T	0.00	0.04	-0.01	-0.05	.962
GALNT13 (rs10196189)	G/G-G/A vs. A/A	-0.12	0.04	-0.33	-2.92	**.004***

HIF1A (rs11549465)	T/T-T/C vs. C/C	-0.08	0.05	-0.22	-1.73	.086
HSD17B14 (rs7247312)	A/A vs. A/G-G/G	0.00	0.05	0.00	0.02	.981
IGF1 (rs35767)	A/A-A/G vs. G/G	-0.10	0.04	-0.28	-2.40	**.018***
IGF2 (rs680)	T/T-T/C vs. C/C	-0.05	0.04	-0.15	-1.35	.179
IL6 (rs1800795)	G/G vs. G/C-C/C	0.01	0.04	0.04	0.37	.712
NOS3 (rs2070744)	C/C-C/T vs. T/T	-0.02	0.04	-0.04	-0.38	.703
PPARA (rs4253778)	C/C-C/G vs. G/G	0.03	0.04	0.08	0.72	.475
PPARG (rs1801282)	G/G-G/C vs. C/C	-0.06	0.05	-0.17	-1.22	.224
SLC16A1 (rs1049434)	T/T vs. T/A-A/A	-0.03	0.04	-0.09	-0.73	.466
SOD2 (rs4880)	G/G vs. A/A-A/G	-0.02	0.04	-0.07	-0.57	.572
TRHR (rs7832552)	C/C vs. C/T-T/T	-0.07	0.04	-0.21	-1.92	.057
UCP2 (rs660339)	G/G vs. G/A-A/A	0.03	0.04	0.08	0.65	.517
5-0-5 agility						
ACE (rs4341)	C/C-C/G vs. G/G	0.01	0.03	0.09	0.46	.649
ACTN3 (rs1815739)	C/C vs. C/T-T/T	-0.01	0.02	-0.06	-0.31	.758
ADBR2 (rs1042714)	C/C vs. C/G-G/G	-0.02	0.03	-0.12	-0.58	.561
AGT (rs699)	G/G-G/A vs. A/A	0.01	0.02	0.06	0.30	.768
AMPD1 (rs17602729)	G/A-A/A vs. G/G	-0.03	0.03	-0.25	-1.05	.298
BDNF (rs6265)	T/T-T/C vs. C/C	-0.02	0.02	-0.12	-0.63	.528
CKM (rs8111989)	T/T vs. T/C-C/C	0.02	0.02	0.18	1.00	.321
CPNE5 (rs3213537)	T/T-T/C vs. C/C	0.00	0.03	0.01	0.06	.950
FTO (rs9939609)	T/A-A/A vs. T/T	0.03	0.03	0.23	1.14	.256

GALNT13 (rs10196189)	G/G-G/A vs. A/A	0.03	0.03	0.22	1.04	.302
HIF1A (rs11549465)	T/T-T/C vs. C/C	0.00	0.03	0.04	0.16	.877
HSD17B14 (rs7247312)	A/A vs. A/G-G/G	0.04	0.03	0.29	1.26	.210
IGF1 (rs35767)	A/A-A/G vs. G/G	0.02	0.02	0.14	0.70	.488
IGF2 (rs680)	T/T-T/C vs. C/C	-0.01	0.02	-0.06	-0.33	.744
IL6 (rs1800795)	G/G vs. G/C-C/C	0.02	0.02	0.16	0.82	.412
NOS3 (rs2070744)	C/C-C/T vs. T/T	0.00	0.02	-0.02	-0.09	.927
PPARA (rs4253778)	C/C-C/G vs. G/G	0.02	0.02	0.15	0.78	.438
PPARG (rs1801282)	G/G-G/C vs. C/C	-0.05	0.03	-0.37	-1.58	.117
SLC16A1 (rs1049434)	T/T vs. T/A-A/A	0.01	0.02	0.10	0.51	.610
SOD2 (rs4880)	G/G vs. A/A-A/G	-0.02	0.03	-0.14	-0.64	.526
TRHR (rs7832552)	C/C vs. C/T-T/T	0.01	0.02	0.08	0.43	.671
UCP2 (rs660339)	G/G vs. G/A-A/A	-0.01	0.03	-0.07	-0.31	.755
CMJ						
ACE (rs4341)	C/C-C/G vs. G/G	0.80	1.79	0.10	0.45	.656
ACTN3 (rs1815739)	C/C vs. C/T-T/T	-0.08	1.62	-0.01	-0.05	.960
ADBR2 (rs1042714)	C/C vs. C/G-G/G	-1.16	1.82	-0.15	-0.64	.527
AGT (rs699)	G/G-G/A vs. A/A	1.33	1.69	0.17	0.79	.431
AMPD1 (rs17602729)	G/A-A/A vs. G/G	-4.77	1.95	-0.62	-2.45	**.016***
BDNF (rs6265)	T/T-T/C vs. C/C	0.67	1.58	0.09	0.43	.671
CKM (rs8111989)	T/T vs. T/C-C/C	0.72	1.56	0.09	0.46	.648
CPNE5 (rs3213537)	T/T-T/C vs. C/C	-1.34	1.69	-0.18	-0.79	.430

FTO (rs9939609)	T/A-A/A vs. T/T	0.39	1.72	0.05	0.23	.821
GALNT13 (rs10196189)	G/G-G/A vs. A/A	1.32	1.71	0.17	0.77	.442
HIF1A (rs11549465)	T/T-T/C vs. C/C	2.45	1.86	0.32	1.31	.193
HSD17B14 (rs7247312)	A/A vs. A/G-G/G	-1.15	2.03	-0.15	-0.57	.573
IGF1 (rs35767)	A/A-A/G vs. G/G	0.31	1.62	0.04	0.19	.851
IGF2 (rs680)	T/T-T/C vs. C/C	-1.67	1.56	-0.22	-1.07	.289
IL6 (rs1800795)	G/G vs. G/C-C/C	6.42	1.53	0.84	4.21	**< .001****
NOS3 (rs2070744)	C/C-C/T vs. T/T	1.68	1.60	0.22	1.05	.298
PPARA (rs4253778)	C/C-C/G vs. G/G	-0.42	1.64	-0.06	-0.26	.799
PPARG (rs1801282)	G/G-G/C vs. C/C	2.12	1.83	0.28	1.16	.251
SLC16A1 (rs1049434)	T/T vs. T/A-A/A	2.76	1.57	0.36	1.76	.083
SOD2 (rs4880)	G/G vs. A/A-A/G	-3.56	1.82	-0.47	-1.95	.054
TRHR (rs7832552)	C/C vs. C/T-T/T	0.81	1.55	0.11	0.52	.602
UCP2 (rs660339)	G/G vs. G/A-A/A	3.37	1.81	0.44	1.86	.066

Note. Bold values and * highlight statistical significance at $p < .05$. Bold values and ** highlight statistical significance following false discovery rate correction at 0.05. CMJ = countermovement jump; *B* = unstandardised beta; *SE B* = standard error; β = standardised beta.

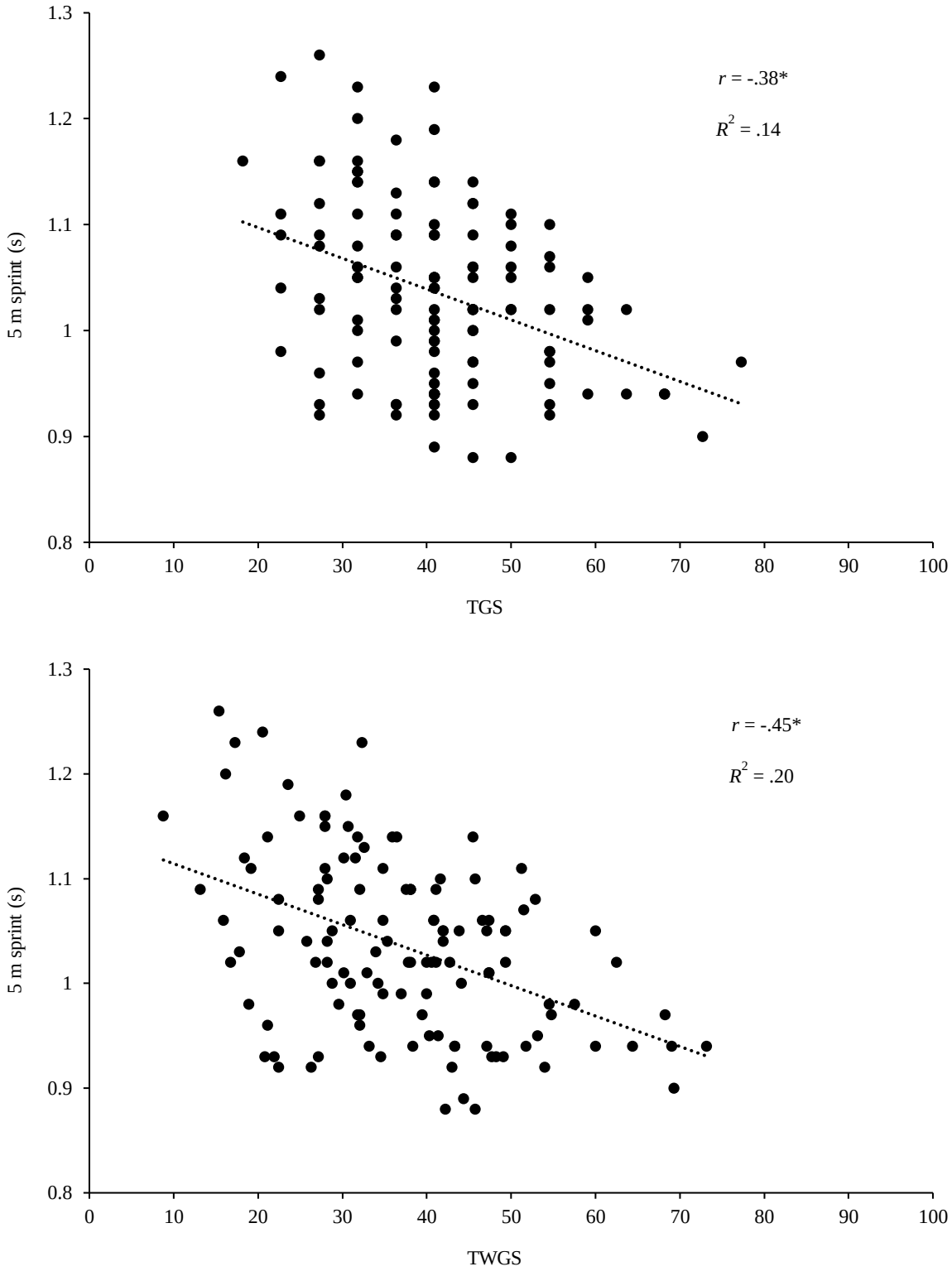


Figure 8.1. 5 m sprint total genotype score (TGS) and total weighted genotype score (TWGS) correlations. * Statistically significant at $p < .05$.

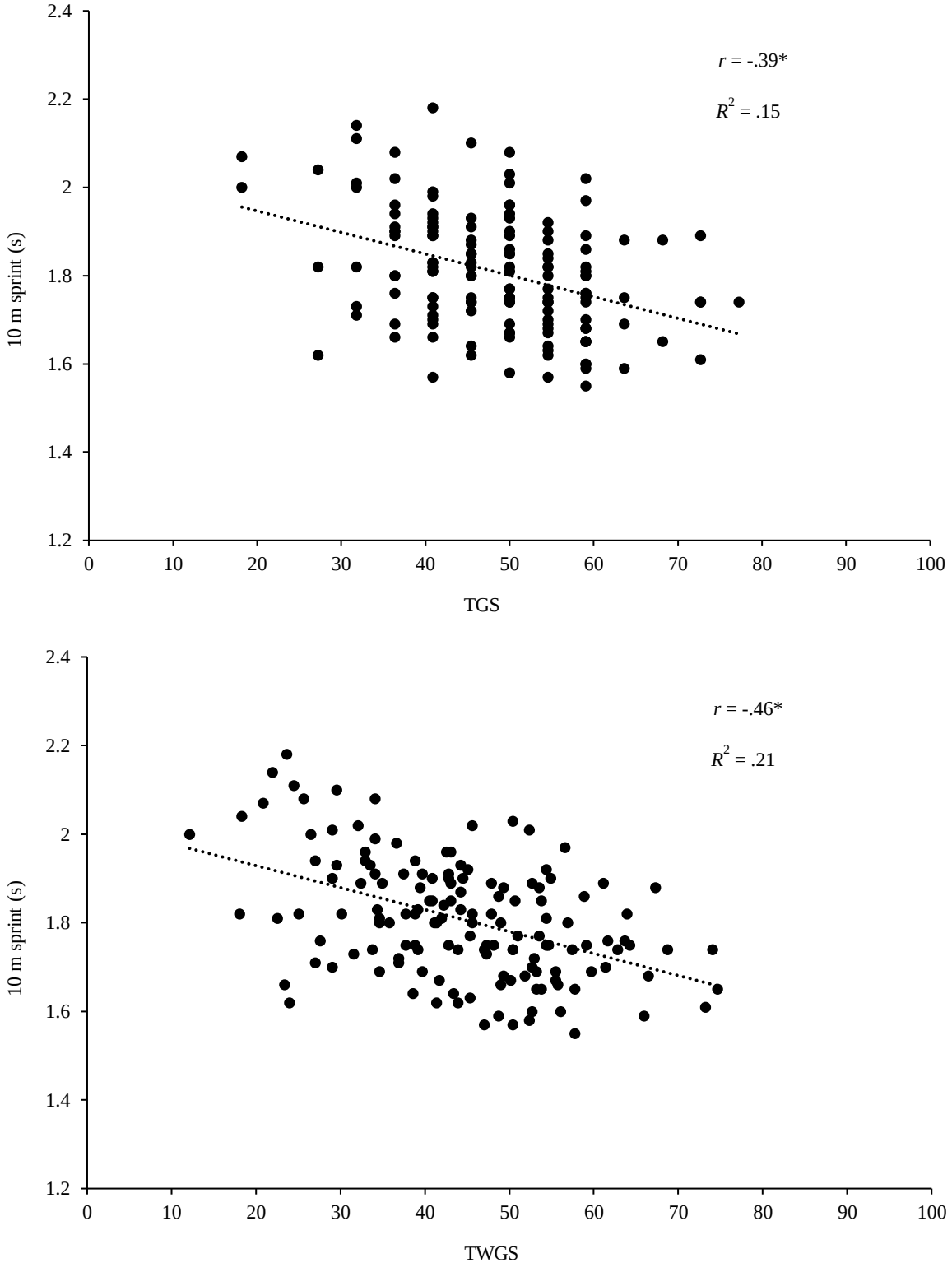


Figure 8.2. 10 m sprint total genotype score (TGS) and total weighted genotype score (TWGS) correlations. * Statistically significant at $p < .05$.

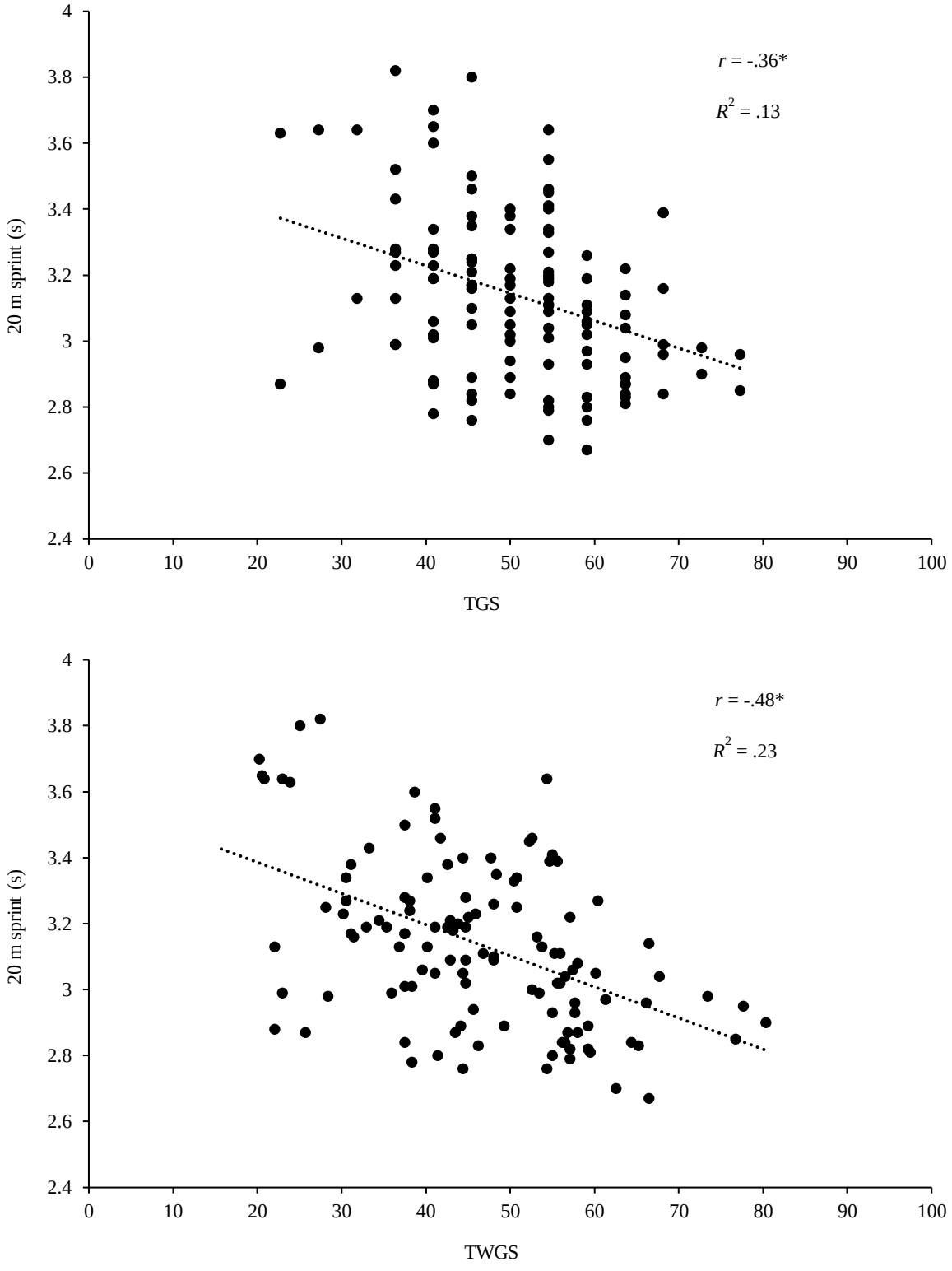


Figure 8.3. 20 m sprint total genotype score (TGS) and total weighted genotype score (TWGS) correlations. * Statistically significant at $p < .05$.

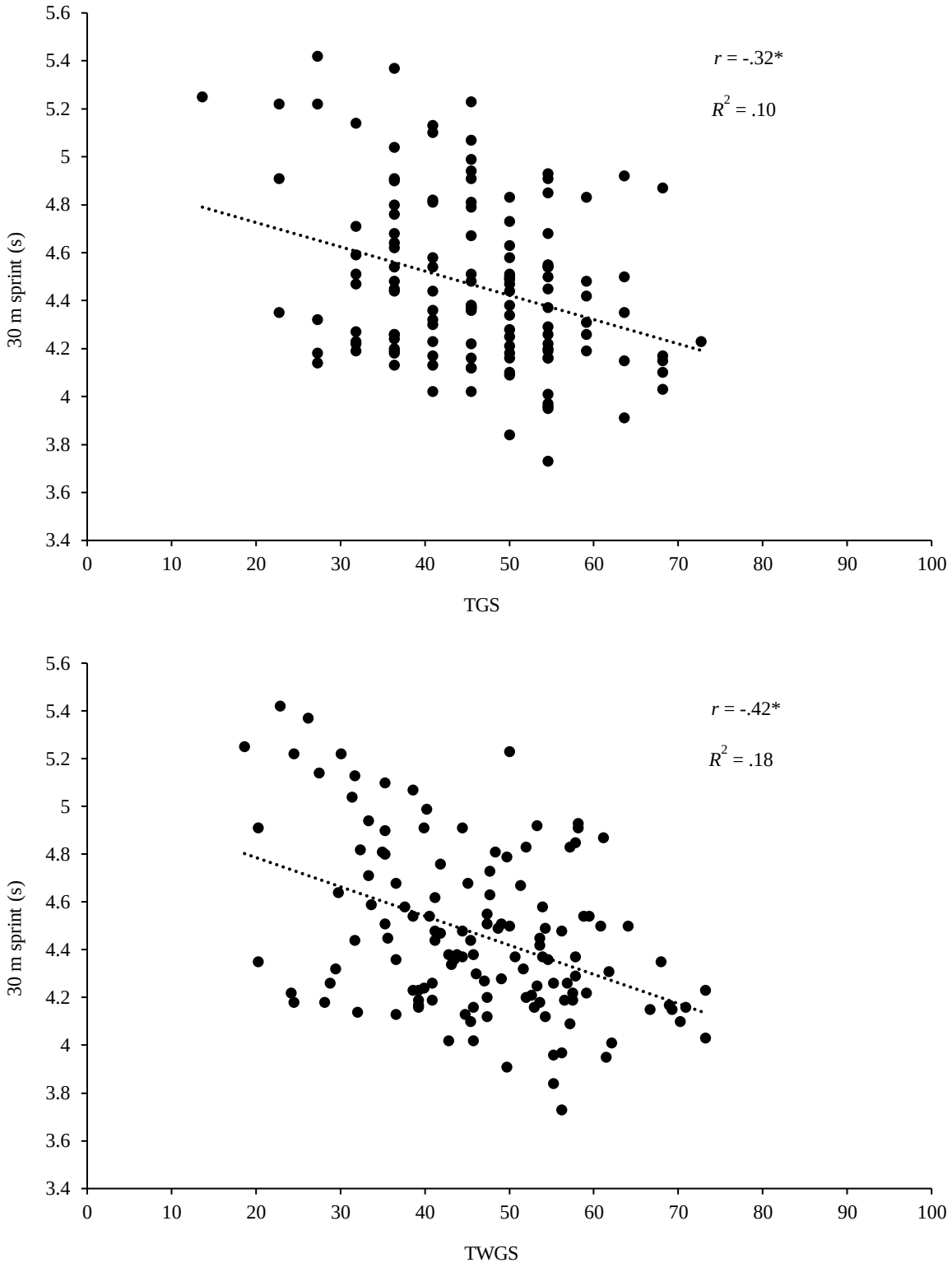


Figure 8.4. 30 m sprint total genotype score (TGS) and total weighted genotype score (TWGS) correlations. * Statistically significant at $p < .05$.

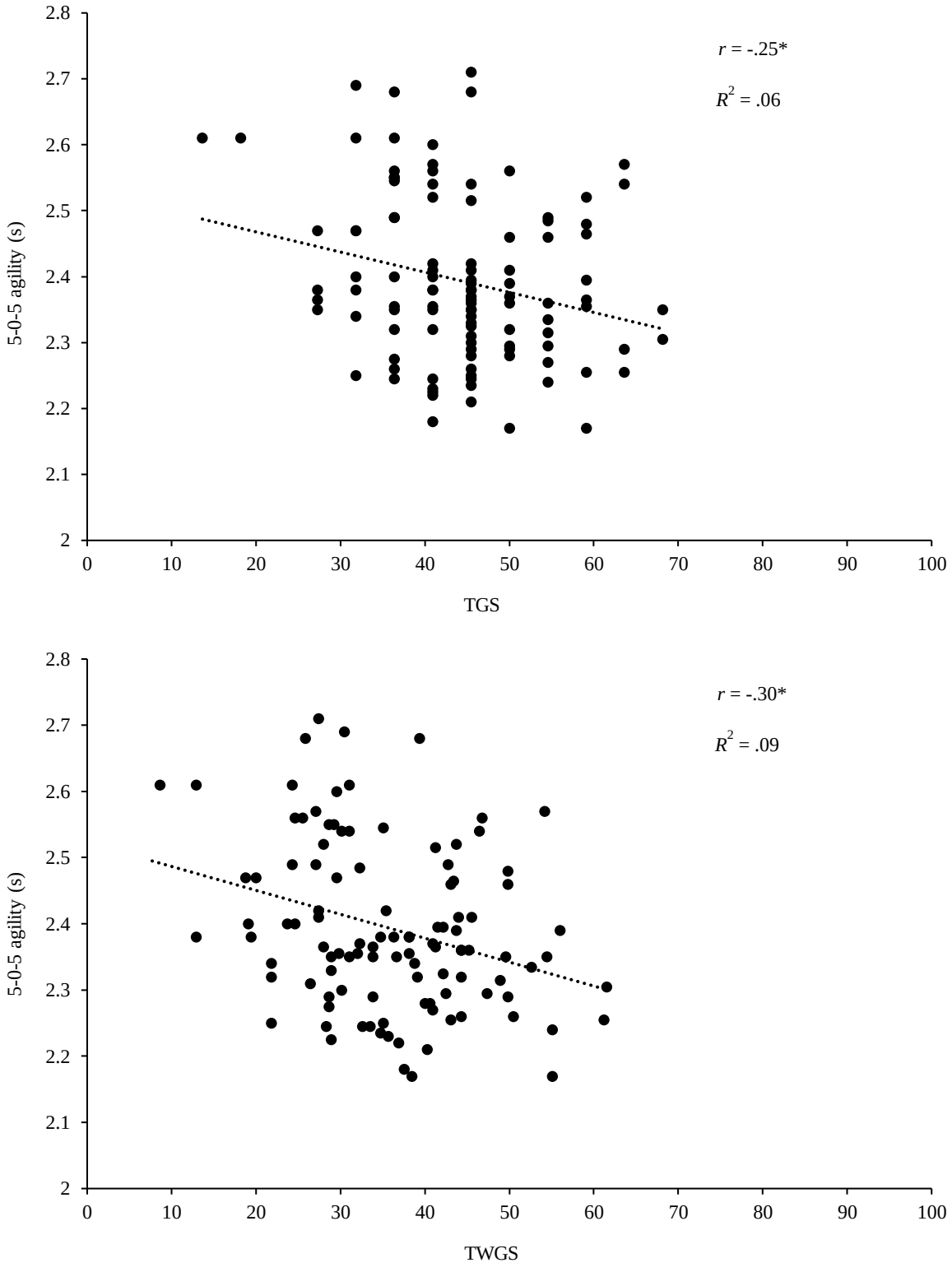


Figure 8.5. 5-0-5 agility total genotype score (TGS) and total weighted genotype score (TWGS) correlations. * Statistically significant at $p < .05$.

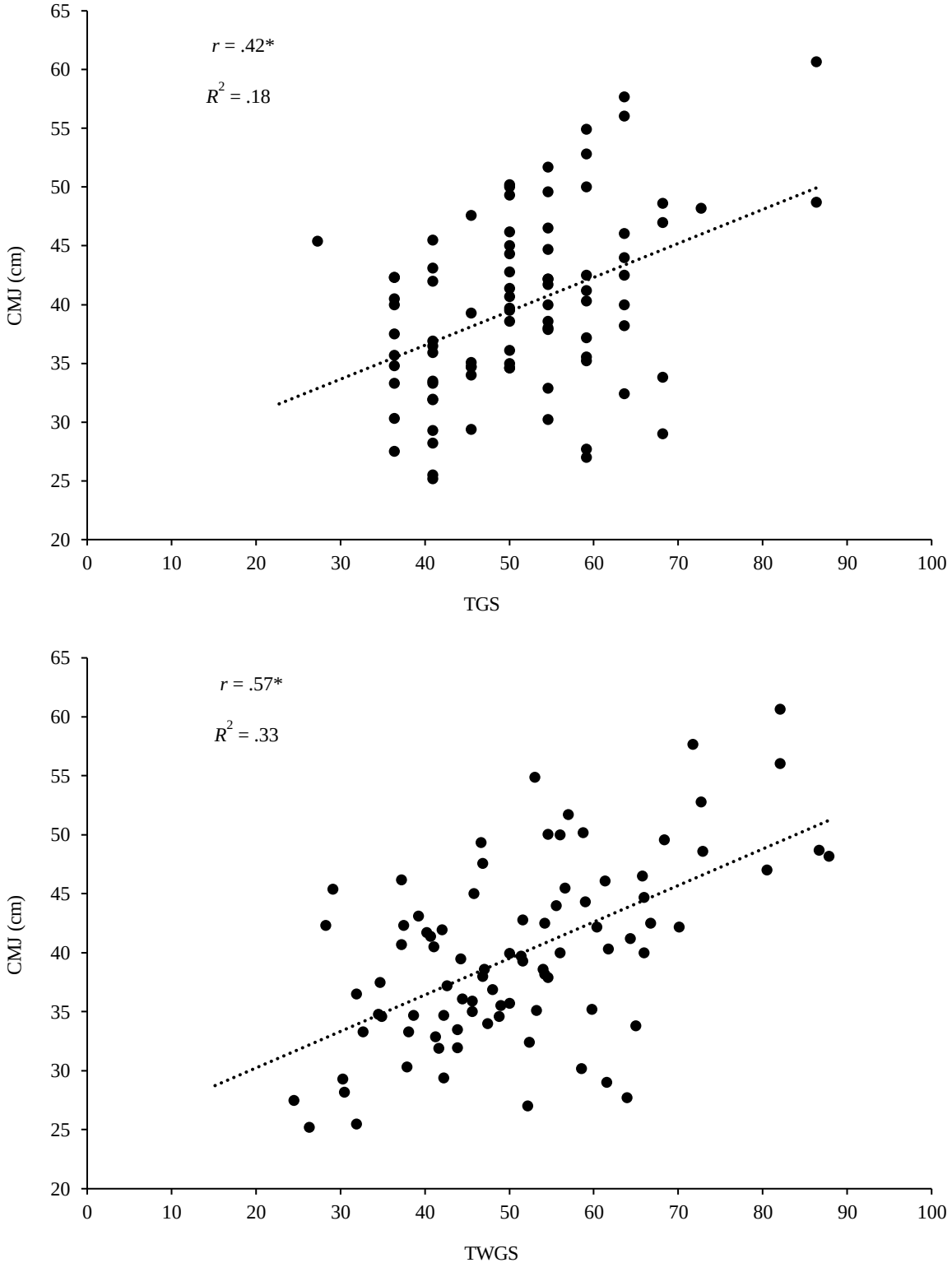


Figure 8.6. Countermovement jump (CMJ) total genotype score (TGS) and total weighted genotype score (TWGS) correlations. * Statistically significant at $p < .05$.

8.4 Discussion

The purpose of this study was to examine the association of twenty-two SNPs, both individually and collectively, with acceleration, COD, jump height, and speed in academy football players. Before FDR correction this study showed associations between ten SNPs in isolation with 5 m, 10 m, 20 m, and 30 m sprint times as well as CMJ performance. Namely, *ACE* (rs4341), *ACTN3* (rs1815739), *AMPD1* (rs17602729), *CKM* (rs8111989), *GALNT13* (rs10196189), *HIF1A* (rs11549465), *IGF1* (rs35767), *IGF2* (rs680), *IL6* (rs1800795), and *SOD2* (rs4880). However, following FDR correction, only the associations between *GALNT13* (rs10196189) and *IL6* (rs1800795) remained significant. Moreover, the T(W)GS models derived from all twenty-two SNPs demonstrated that the advantageous genotypes of each SNP have small but additive effects on all performance assessments. To the author's knowledge, this is the most comprehensive polygenic profile within football genomics that highlights the relationship between genetic variation and physical performance in football.

The *GALNT13* (rs10196189) SNP was associated with performance in the largest number of physical performance assessments and had the strongest individual effect of any SNP. More specifically, the G allele was associated with faster 5 m, 10 m, and 20 m sprint times, compared to the A/A genotype. These findings correspond with Wang et al. (2014), who initially discovered a significant overrepresentation of the G allele in a group of high performing African American, Jamaican, and Japanese sprinters compared to controls, following three genome-wide association studies (GWASs) and subsequent meta-analysis. To the authors' knowledge, this is the first validation of the potential influence of *GALNT13* (rs10196189) on sprinting performance, and the first to explore associations with specific quantitative traits. The *GALNT13* gene encodes the GALNT13 protein, which is highly

expressed in the brain, B cells, kidney, and liver as well as potentially being involved in metabolism and energy pathways (Maciejewska-Skrendo et al., 2019). The GALNT13 protein is a member of the GALNT family, which initiate O-linked glycosylation of mucins by the initial transfer of N-acetyl- galactosamine with an alpha-linkage to a serine or threonine residue, and thus catalyses the initial reaction in O-linked oligosaccharide biosynthesis (Ten Hagen et al., 2003). The rs10196189 SNP is an intron variant, however, currently there are no data regarding the mechanism(s) underpinning the allelic associations. As such, mechanistic studies are required to elucidate the associations of *GALNT13* (rs10196189) with power-orientated phenotypes.

The G/G genotype of *IL6* (rs1800795) was associated with a higher CMJ, compared to the C allele. These findings align with those of Eider et al. (2013) and Ruiz et al. (2010), who both reported that the G allele and G/G genotype were significantly overrepresented in high-performing Polish and Spanish Caucasian power athletes (i.e., jumpers, short-distance swimmers, sprinters, throwers, and weightlifters) compared to controls. Furthermore, Pickering et al. (2019) reported that in a group of British Caucasian youth football players, G allele carriers were significantly faster in 5 m and 20 m sprint tests. As such, the results of this study provide further quantitative data that supports the proposed role of *IL6* (rs1800795) on power-orientated phenotypes. The *IL6* gene encodes for interleukin-6 (IL-6), which is a pleiotropic cytokine involved in several biological processes, including glucose homeostasis, hypertrophic muscle growth, immune function, and muscle damage repair (Febbraio & Pedersen, 2002; Serrano et al., 2008; Yamin et al., 2008). The *IL6* (rs1800795) SNP alters transcriptional response and the subsequent plasma levels of IL-6, with the G allele associated with higher levels (Fishman et al., 1998). Increased IL-6 activity reduces muscle inflammation by

stimulating and inhibiting the production of anti-inflammatory and proinflammatory cytokines, respectively (Petersen and Pedersen, 2005). In addition, an increase in IL-6 released by skeletal muscle fibres following acute exercise enhances adenosine monophosphate-activated protein kinase activity, which can improve glucose utilisation and sustain muscle energy demands (Carey et al., 2006). Moreover, the C allele has been associated with higher creatine kinase activity following eccentric exercise (Yamin et al., 2008). During powerful muscle contractions the G allele may therefore protect skeletal muscle, aid in repair, and promote beneficial adaptations during power training, which may explain why G/G homozygotes jumped higher than C allele carriers.

Eight other SNPs were nominally associated (i.e., before FDR correction) with at least one of the power-orientated phenotypes assessed. The advantageous genotypes/alleles of five SNP associations (i.e., *ACTN3* rs1815739 C/C, *AMPD1* rs17602729 G/G, *HIF1A* rs11549465 C/C, *IGF1* rs35767 A allele, and *SOD2* rs4880 G/G) correspond with most of the previous research on power/strength athletes (see Ahemetov et al., 2021 for a recent review). As such, although these SNPs did not pass the statistical threshold in this study, the current evidence base provides support for the direction of their associations. However, the advantageous genotypes/alleles of the remaining three SNP associations (i.e., *ACE* rs4341 C allele, *CKM* rs8111989 T/T, and *IGF2* rs680 T allele) do not align with most of the previous research on power/strength athletes. This suggests the direction of these associations may be less trustworthy and have lower external validity. Although, it should be noted that previous research is predominately comprised of athlete status studies that use case-control designs to compare the genotype frequencies of athletes to non-athletes or athletes of endurance-orientated sports. As such, it is unclear which functional phenotypes underpinning athlete status

are being influenced by the allelic changes via these SNPs. Further research on the relationship between genotype frequency and physiological measures with larger sample sizes is required to better understand how genetic variation affects performance and consequently athlete status.

The TGSs were associated with every physical performance test. This suggests acceleration, COD, jump height, and speed are influenced by the combination of several genetic variants with small but additive effects. These results are not surprising, as they are in accordance with the findings of previous polygenic case-control investigations on power-orientated athletes. For instance, Ruiz et al. (2010) originally reported a TGS derived from six SNPs (five of which were used in this study) differentiated high-performing Spanish Caucasian jumpers and sprinters from distance runners, road cyclists, and non-athletic controls. Whilst, more recently, Moreland et al. (2020) found high-performing Russian Caucasian weightlifters and powerlifters had a significantly higher TGS comprising 28 SNPs (three of which were used in this study) than controls. The findings of other cross-sectional studies investigating quantitative power-orientated phenotypes are also comparable. For example, Murtagh et al. (2020) revealed a TGS derived from four SNPs (two of which were used in this study) was associated with the performance of youth footballers in England and Uruguay in 10 m and 20 m sprints, horizontal and vertical CMJs, and a modified 5-0-5 agility test. Moreover, Petr et al. (2021) found a TGS comprising seven polymorphisms (five of which were used in this study) was positively correlated with the CMJ and squat jump height of Caucasian professional football players in the Czech Republic.

The TWGSs were also associated with acceleration, COD, jump height, and speed, but consistently displayed stronger relationships than the TGSs across all physical performance tests. This suggests that whilst the SNPs have small additive effects on every power-orientated

phenotype, each advantageous allele of a given SNP has a different degree of influence as suggested previously (see Massidda et al., 2014). These TWGS associations correspond and expand on the work of Varillas Delgado et al. (2020), who found a TWGS derived from four SNPs differentiated high performing Spanish Caucasian road cyclists and endurance runners from sedentary controls. Furthermore, TWGSs displaying superior relationships than TGSs with quantitative physical performance phenotypes is also congruent with previous research using similar approaches. For instance, Charlier et al. (2017) reported a TWGS comprising nine polymorphisms was a better predictor of isometric strength, isokinetic strength, and ballistic movement speed in Flemish Caucasian males. Whereas Massidda et al. (2014) revealed a TWGS derived from three polymorphisms (two of which were used in this study) explained more CMJ (10%) and squat jump (15%) variance in professional football players in Italy. In this study, the TWGSs explained 3-15% more of the variance than the TGSs across all performance tests.

The associations of the T(W)GSs reinforce the utility of polygenic profiles to identify higher performing players in power-orientated performance assessments. However, the findings also demonstrate that subtle distinctions in the phenotypic characteristics of different power measurements may alter allelic associations and consequent relationships with polygenic profiles. Therefore, the specific characteristics underpinning a phenotype require careful evaluation before utilising a polygenic profile to optimise its practicality. This is because each gene has very specific biological functions that can affect individual phenotypes differently depending on their underlying physiology. The largest amount of variance explained by the polygenic profile in this study was in CMJ performance (33%) and the least amount of variance explained was in 5-0-5 agility performance (9%). As such, this polygenic

profile may be best utilised for jump height assessments and development, however, further replicatory research is required. Given the importance of physical performance in football, successful independent replication in large homogenous cohorts may allow future implementation of more individualised physical intervention programmes during athlete development.

Whilst this study does present several significant findings, it is important to acknowledge it does have some limitations. First, although chronological age was controlled for during analysis, there are many other factors that influence performance in power assessments that could not be controlled for. For instance, recent research has shown that not only does maturation status influence physiological capacities, but it is also associated with distinct genetic profiles in youth football players (Murtagh et al., 2020). Capturing and adjusting for maturation status as well as chronological age and other confounding variables may provide greater context to findings. Second, the participants in this study differed in ethnic origin which may have introduced population stratification issues (i.e., systematic differences in allele frequencies between subpopulations). However, participants were included from multiple academies regardless of ethnicity to more accurately reflect academy player profiles in England and raise external validity. Third, the sample size was relatively small despite recruiting participants from four academies, as higher-performing athletic populations are generally small by nature. Although, the number of participants in this study (N = 149) is considerably higher than the median sample size (N = 60) of eighty previous football genetic studies (see McAuley, Hughes, et al., 2021a for a review). Building this research base with studies using transparent methodologies is important so they can contribute to research

synthesis approaches in the future to draw more valid and reliable conclusions (McAuley, Baker, et al., 2022).

8.5 Conclusion

This study has shown that inter-individual genetic variation is associated with acceleration, COD, jump height, and speed in youth football players. These findings suggest that *GALNT13* (rs10196189) and *IL6* (rs1800795) may be significant predictors of 5 m, 10 m, and 20 m sprint times and CMJ height, respectively. However, the polygenic models derived from all twenty-two SNPs were associated with all physical performance assessments. As such, all the SNPs included in this study may add value to a genetic profile tool that could facilitate more individualised physical training programmes during athlete development. Although, before practical applications can be contemplated, the external validity of these findings should be assessed via replication studies in larger independent and homogenous football cohorts. Moreover, the identification of more SNPs and other genetic variants relevant to performance is necessary to improve the sensitivity and specificity of polygenic profiles.

9 Genetic variations between youth and professional development phase English academy football players

McAuley, A. B. T., Hughes, D. C., Tsaprouni, L. G., Varley, I., Suraci, B., Baker, J., Herbert, A. J., & Kelly, A. L. (2022c). Genetic variations between youth and professional development phase English academy football players. *Genes, 13*(11), 2001. https://doi.org/10.3390/genes13112001

Abstract

The purpose of this study was to examine differences in the genotype frequency distribution of thirty-three single nucleotide polymorphisms (SNPs) between youth development phase (YDP) and professional development phase (PDP) academy football players. One hundred and sixty-six male football players from two Category 1 and Category 3 English academies were examined within their specific age phase: YDP (n = 92; aged 13.84 ± 1.63 years) and PDP (n = 74; aged 18.09 ± 1.51 years). Fisher's exact tests were used to compare individual genotype frequencies, whereas unweighted and weighted total genotype scores (TGS; TWGS) were computed to assess differences in polygenic profiles. In isolation, the *IL6* (rs1800795) G allele was overrepresented in PDP players (90.5%) compared to YDP players (77.2%; p = .023), whereby PDP players had nearly three times the odds of possessing a G allele (OR = 2.83, 95% CI: 1.13-7.09). Furthermore, the TGS (p = .001) and TWGS (p < .001) were effective in distinguishing YDP and PDP players (AUC = 0.643-0.694), with PDP players exhibiting an overall more power-orientated polygenic profile. If validated in larger independent youth

football cohorts, these findings may have important implications for future studies examining genetic associations in youth football.

9.1 Introduction

The process of athlete development and ultimately reaching senior professional status in a sport such as football (soccer) is both dynamic and multifactorial (McAuley, Baker, et al., 2021). Indeed, task constraints (e.g., the value of deliberate practice and deliberate play, or the importance of early engagement), performer constraints (e.g., differences between skill levels on anthropometric/physiological factors, psychological characteristics, and technical or tactical skill), and environmental constraints (e.g., the influence of birth-place and relative age effects and/or socio-cultural influences) have all been associated with the performance of youth football players and their potential to achieve adult success (Sarmento et al., 2018). Despite being heavily researched, the extent to which each of these elements impact performance and affects the likelihood of achieving senior professional status in football remains unclear (Williams et al., 2020).

The failure to clearly identify a set of variables that uniformly predicts performance levels is, in part, due to methodological issues identified throughout talent identification and development research in football (see Bergkamp et al., 2019). Prospective and longitudinal analyses in youth football have also revealed that specific performer characteristics may be more important at different time-points throughout development (see Abarghoueinejad et al., 2021 for a review). For instance, when comparing English academy football players of different age groups (i.e., under-9 to under-11 vs. under-12 to under-16) Kelly and colleagues showed important differences in physical characteristics and decision-making (Kelly et al., 2021a), technical skill (Kelly et al., 2020), as well as differentiating those who ‘play-up’ an age group (Kelly et al., 2021b). From a longitudinal perspective, Saward et al. (2020) performed a ten-year prospective investigation of 2,875 male youth football players (aged 8-19 years) from 16

English academies, and found that future professionals only began to significantly outperform their non-professional counterparts in vertical countermovement jump (CMJ; > 0.6 cm) and slalom agility performance (< 0.03 s) at the age of 12 years. Moreover, at the age of 18 years these differences were significantly greater (i.e., > 1.7 cm and < 0.14 s, respectively), and thus had superior prognostic power.

Although under-researched within a football context, inter-individual genetic variation also appears to influence performance and development in football (see McAuley, Hughes et al., 2021a for a review). Moderate to high heritability estimates (i.e., 30-80%) have been reported for anthropometric (e.g., height and skeletal muscle mass = 80%), physiological (e.g., strength and power = 52%), psychological (e.g., personality dimensions and mental toughness = 50%), and technical (e.g., motor control and motor learning = 70%) factors (Horsburgh et al., 2009; Livshits et al., 2016; Missitzi et al., 2013; Silventoinen et al., 2008; Zempo et al., 2017). Moreover, there have been sizeable heritability estimates reported for specific injuries such as anterior cruciate ligament rupture (69%) and overall athlete status (66%) (De Moor et al., 2007; Magnusson et al., 2021).

Recent studies have also begun to explore which specific genetic variants may explain some of the genetic influence on performance and development in football (e.g., McAuley, Hughes et al., 2022a, 2022b; 2022e; Murtagh et al., 2020). However, the majority of genetic research in football is comprised of case-control athlete status designs, which have had limited success (McAuley, Hughes, et al., 2021a, 2021b). Given that the importance of specific characteristics during development in football appears to alter depending on age, the genetic profiles of youth players may also differ between distinct age groups. Indeed, recent research on maturation showed that the genotype frequency distributions of four genetic variants (i.e.,

ACTN3 rs1815739, *AGT* rs699, *PPARA* rs4253778, and *NOS3* rs2070744) were significantly different between pre- (aged 10.6 ± 1.4 years) and post- (aged 16.8 ± 2.3 years) peak height velocity academy football players (Murtagh et al., 2020).

In England, the structure of football academies is governed by the Elite Player Performance Plan (EPPP; Premier League, 2013), with age groups divided into three development phases, two of which include the youth development phase (YDP; under-12 to under-16) and professional development phase (PDP; under-17 to under-23). The purpose of this study was to examine differences in the genotype frequency distribution, both individually and collectively, of thirty-three single nucleotide polymorphisms (SNPs) between YDP and PDP academy football players. Such information may have important implications for future studies examining genetic associations in youth football, as well as advance methodological approaches within this field of research.

9.2 Methods

9.2.1 Participants

One hundred and sixty-six male football players from two Category 1 and Category 3 English academies participated within their specific age phase: YDP (n = 92; aged 13.84 ± 1.63 years) and PDP (n = 74; aged 18.09 ± 1.51 years). Informed assent from all players, consent from parents/guardians, and gatekeeper consent from each academy were collected prior to the commencement of the study. All experimental procedures were conducted in accordance with the guidelines in the Declaration of Helsinki and ethical approval was granted by the corresponding author's institutional Ethics Committee. This study was conducted in accordance with the recommendations for reporting the results of genetic association studies

defined by the STrengthening the REporting of Genetic Association studies (STREGA) Statement.

9.2.2 Genetic procedures

9.2.2.1 Genotyping

Saliva was collected from players via sterile, self-administered buccal swabs, following a minimum of 30-minutes since food or drink ingestion. Within 36-hours, saliva samples were sent to AKESOgen, Inc. (Peachtree Corners, GA, USA) for DNA extraction. Using Qiagen chemistry, DNA was extracted on an automated Kingfisher FLEX instrument (Thermo Fisher Scientific, Waltham, MA, US). To measure the quality and quantity of extracted DNA, PicoGreen and Nanodrop measurements were taken. Input to the custom testing array occurs at 200 ng in 20 μL. Amplification, fragmentation, and resuspension were performed using Biomek FXP. GeneTitan instrumentation (Thermo Fisher Scientific, Waltham, MA, US) was used to stain and scan the arrays, with hybridisation performed in a Binder oven at 48-degrees for 24-hours, following the Affymetrix Axiom high throughput 2.0 protocol. Data analysis was then performed using a raw CEL file data input into the Affymetrix Axiom Analysis Suite (Affymetrix, Santa Clara, CA, US). Procedures were in accordance with Pickering et al. (2019).

9.2.2.2 Polymorphism selection

Thirty-three SNPs were selected based on their proposed biological function and relevant associations with physiological, psychological, technical, and injury phenotypes in previous studies with academy football players (Hall et al., 2022; McAuley, Hughes, et al., 2022a, 2022b; Murtagh et al., 2020; Pickering et al., 2019) (see **Table 9.1**). Gene names and symbols are in accordance with those officially approved by the Human Gene Nomenclature Committee (HGNC; https://www.genenames.org). Standard genomic quality control (QC) procedures and

thresholds were applied when selecting polymorphisms: SNP call rate (>95), sample call rate (>95), fisher's linear discriminant (>3.6), and minor allele frequency (>0.05).

9.2.2.3 Total genotype score

Unweighted and weighted total genotype scores (TGS; TWGS) were calculated to assess the differences in polygenic profiles between YDP and PDP players. Both TGSs and TWGSs have demonstrated sufficient discriminatory power in previous sport genomic research (Varillas Delgado et al., 2020; Williams & Folland, 2008). To generate both the TGS and TWGS, each genotype of a respective SNP initially received a score between 0-2 using a data-driven approach based on the observed genotype associations with PDP status. Genotypes of dominant (AA vs. Aa-aa) and recessive (AA-Aa vs. aa) models were assigned a score of two (i.e., associated genotype[s]) or zero (i.e., alternate genotype[s]), whereas genotypes of co-dominant models (AA vs. Aa vs. aa) were assigned three scores (i.e., homozygous-associated genotypes received a score of two, the heterozygote received a score of one, and the alternate homozygous genotype received a score of zero).

For the TGS, the original procedure of Williams and Folland (2008) was followed. Genotype scores (GS) were summed and transformed into a 0-100 scale by dividing the total score by the maximum possible score and multiplying by 100.

$$TGS = (combined\text{-}GS / maximum\text{-}GS) * 100$$

For the TWGS, a similar procedure to Varillas Delgado et al. (2020) was used. Each GS was multiplied by the beta coefficients of each SNP following multiple regression to create weighted genotype scores (WGS). The WGSs were then summed and transformed into a 0-100 scale by dividing the total score by the maximum possible score and multiplying by 100.

TWGS = (combined-WGS / maximum-WGS) * 100

9.2.3 Data analysis

Data were analysed using Jamovi version 1.8.1 and IBM SPSS version 25. Fisher's exact tests were used to test SNPs for adherence with Hardy-Weinberg equilibrium (HWE) and to compare genotype frequencies between YDP and PDP players. Akaike information criterion (AIC) was used to select which genetic model (i.e., co-dominant, dominant, recessive) best fit the data and would be subjected to hypothesis testing. However, if MAF ≤ 0.25 a dominant model was utilised to retain statistical power (Murtagh et al., 2020). An independent t-test was used to assess differences in TGS and TWGS between YDP and PDP players. Additionally, receiver operating characteristic (ROC) curves and area under the curve (AUC) were used to evaluate the discriminatory power of the TGS and TWGS to distinguish YDP and PDP players with threshold values of: >0.5-0.7 = poor, >0.7-0.8 = acceptable, >0.8-0.9 = excellent, and >0.9 = outstanding (Hosmer et al., 2013). Odds ratios (OR) and 95% confidence intervals (CI) were also calculated to estimate the effect size of individual genotypes and polygenic models (split into equal thirds using tertiles). Statistical significance was set at $p < .05$.

Table 9.1. Gene and single nucleotide polymorphism (SNP) information.

Gene	Symbol	Chr	Function	SNP	Consequence	Associations	MAF
5-hydroxytryptamine receptor 2A	*HTR2A*	13q14.2	Encodes one of the receptors for serotonin, a neurotransmitter with an important role in the brain reward system as well as cellular development and differentiation.	rs6311	Intron variant C > T	Big Five personality traits and novelty seeking.	T = 0.44
Actinin alpha 3	*ACTN3*	11q13.2	Primarily expressed in skeletal muscle and functions as a structural component of sarcomeric Z line.	rs1815739	Nonsense variant C > T (Arg > Ter)	Protein deficiency, percentage of muscle fibre types, strength, muscle volume, acceleration, speed, and power.	T = 0.43
Adenosine monophosphate deaminase 1	*AMPD1*	1p13.2	Catalyses the deamination of AMP to IMP in skeletal muscle.	rs17602729	Nonsense variant G > A (Gln > Ter)	Enzyme functionality, AMP metabolism, muscle fatigue, weakness and cramping.	A = 0.12
Adrenoceptor beta 2	*ADRB2*	5q32	Expressed in the brain as part of the catecholamine system and acts as a receptor for adrenaline, noradrenaline, and dopamine.	rs1042714	Missense variant G > C (Glu > Gln)	Receptor activity and sensitivity, neurodevelopmental disorders, white matter and cognitive functions.	G = 0.41

Angiotensin I converting enzyme	*ACE*	17q23.3	Catalyses the degradation of the inactive decapeptide angiotensin I, and generates the physiologically active peptide, angiotensin II.	rs4341	Intron variant C > G (Insertion > Deletion)	Serum and tissue activity, blood pressure, percentage of muscle fibre types, strength and muscle volume.	C = 0.43
Angiotensinogen	*AGT*	1q42.2	Expressed in the liver and is cleaved by the enzyme renin in response to lowered blood pressure and produces angiotensin I.	rs699	Missense variant A > G (Met > Thr)	Plasma levels, serum concentration, blood pressure, skeletal muscle hypertrophy and volume.	G = 0.41
Brain derived neurotrophic factor	*BDNF*	11p14.1	Encodes for a neurotrophin essential for neuronal development and survival, synaptic plasticity, and cognitive function.	rs6265	Missense variant C > T (Val > Met)	Motor cortex plasticity, motor learning, neuromuscular activation, horizontal and vertical power.	T = 0.20
Catechol-O-methyltransferase	*COMT*	22q11.21	Catalyses the degradation of catecholamines, including neurotransmitters dopamine, epinephrine, and norepinephrine.	rs4680	Missense variant G > A (Val > Met)	Impulsivity, addiction disorders, extraversion, sensation seeking.	A = 0.50
Catenin alpha 2	*CTNNA2*	2p12	Expressed in the brain with a role in synaptic plasticity, actin filament binding activity, regulation of Arp2/3 complex-mediated actin nucleation, neuron migration and projection development.	rs7600563	Intron variant T > G	Excitement seeking and risk taking.	G = 0.34

Cholinergic receptor muscarinic 2	*CHRM2*	7q33	Muscarinic receptors influence many effects of acetylcholine in the central and peripheral nervous system. The muscarinic cholinergic receptor 2 is involved in mediation of bradycardia and a decrease in cardiac contractility	rs1824024	Intron variant C > A	Big Five personality traits and sensation seeking.	C = 0.29
Copine 5	*CPNE5*	6p21.2	Calcium-dependent membrane-binding proteins may regulate molecular events at the interface of the cell membrane and cytoplasm	rs3213537	Intron variant C > T	Percentage of muscle fibre types, acceleration, speed, and power athlete status.	T = 0.14
Creatine kinase, M-type	*CKM*	19q13.32	A cytoplasmic enzyme involved in energy homeostasis and an important serum marker for myocardial infarction.	rs8111989	500B Downstream variant T > C	Skeletal muscle performance, maximum oxygen uptake, trainability, power and endurance athlete status.	C = 0.30
Dopamine beta-hydroxylase	*DBH*	9q34.2	Catalyses the oxidative hydroxylation of dopamine to norepinephrine modulating executive and motor function.	rs1611115	2KB Upstream variant C > T	Enzymatic activity, metabolism of dopamine and norepinephrine, as well as neurodevelopmental disorders.	T = 0.21

Dopamine receptor D1	*DRD1*	5q35.2	Encodes the D1 subtype of the dopamine receptor, which stimulates adenylyl cyclase and activates cyclic AMP-dependent protein kinases, regulating neuronal growth, development, and behaviour.	rs4532	5 Prime UTR variant C > T	Dopaminergic tone, dopamine neurotransmission, motor learning, neurodevelopmental and neurodegenerative disorders.	C = 0.40
Dopamine receptor D2	*DRD2*	11q23.2	Encodes the D2 subtype of the dopamine receptor, which inhibits adenylyl cyclase activity and regulates neuronal growth, development, and behaviour.	rs1076560	Intron variant C > A	Dopaminergic tone, dopamine neurotransmission, motor learning, neurodevelopmental and neurodegenerative disorders.	A = 0.15
Dopamine receptor D3	*DRD3*	3q13.31	Encodes the D3 subtype of the dopamine receptor, which inhibits adenylyl cyclase activity and is associated with cognitive, emotional, and endocrine functions.	rs6280	Missense variant C > T (Gly > Ser)	Dopaminergic tone, dopamine neurotransmission, motor learning, neurodevelopmental and neurodegenerative disorders.	C = 0.33
Dopamine receptor D4	*DRD4*	11p15.5	Encodes the D4 subtype of the dopamine receptor, which inhibits adenylyl cyclase activity and is associated with autonomic nervous system dysfunction, behaviour and personality traits.	rs1800955	2KB Upstream variant T > C	Novelty seeking, sensation seeking, anxiety and impulsivity.	C = 0.41

FTO alpha-ketoglutarate dependent dioxygenase	*FTO*	16q12.2	Expressed in the hypothalamic nuclei that regulate energy balance.	rs9939609	Intron variant T > A	Fat mass, lean mass, muscle power, percentage of muscle fibre types, and power and endurance athlete status.	A = 0.41
Gamma-aminobutyric acid type A receptor subunit alpha6	*GABRA6*	5q34	GABA is the major inhibitory neurotransmitter in the mammalian brain where it acts at GABA-A receptors, which are ligand-gated chloride channels.	rs3219151	3 Prime UTR variant C > T	Big Five personality traits, panic and major depressive disorder.	C = 0.42
Hydroxysteroid 17-beta dehydrogenase 14	*HSD17B14*	19q13.33	Involved in metabolism of steroids at the C17 position and also of other substrates, such as fatty acids, prostaglandins, and xenobiotics.	rs7247312	Intron variant A > G	Power athlete status, acceleration, and speed.	G = 0.10
Hypoxia inducible factor 1 subunit alpha	*HIF1A*	14q23.2	Regulates cellular and systemic homeostatic response to hypoxia by activating transcription of many genes.	rs11549465	Missense variant C > T (Pro > Ser)	Protein stability, transcriptional activity, percentage of muscle fibre types, and power and endurance athlete status.	T = 0.10
Insulin like growth factor 1	*IGF1*	12q23.2	Protein similar to insulin in function and structure involved in mediating growth and development.	rs35767	Missense variant G > A (Gly > Val)	Circulating IGF1 levels, body composition, power athlete status, and strength.	A = 0.16

Insulin like growth factor 2	*IGF2*	11p15.5	A member of the insulin family of polypeptide growth factors, which are involved in development and growth.	rs680	3 Prime UTR variant C > T	Acceleration, speed, and power athlete status.	T = 0.32
Interleukin 6	*IL6*	7p15.3	A pleiotropic cytokine involved in glucose homeostasis, hypertrophic muscle growth, immune function, and muscle damage repair.	rs1800795	Intron variant G > C	Acceleration, speed, and power and strength athlete status.	C = 0.42
Nitric oxide synthase 3	*NOS3*	7q36.1	A biologic mediator in several processes, potentially stimulating muscle hypertrophy through NO-mediated vasodilatation.	rs2070744	Intron variant C > T	Promoter activity, endothelial nitric oxide synthesis, muscle hypertrophy, acceleration, and power athlete status.	C = 0.44
Oxytocin receptor	*OXTR*	3p25.3	Encodes oxytocin a neuropeptide and hormone implicated in prosocial human social and emotional functioning.	rs2254295	Intron variant C > T	Reward seeking, risk taking, and impulsivity.	C = 0.11
Peroxisome proliferator activated receptor alpha	*PPARA*	22q13.31	Present in skeletal muscle and promotes uptake, utilization, and catabolism of fatty acids by upregulation of other genes.	rs4253778	Intron variant G > C	Percentage of muscle fibre types, strength, power, muscle mass, and power and endurance athlete status.	C = 0.19
Peroxisome proliferator activated receptor gamma	*PPARG*	3p25.2	A central transcriptional regulator of adipogenic and lipogenic programs, insulin sensitivity and glucose homeostasis	rs1801282	Missense variant C > G (Pro > Ala)	Receptor activity, insulin sensitivity, skeletal muscle glucose uptake, cross-sectional area of muscle fibres.	G = 0.12

Polypeptide N-acetylgalactosaminyltransferase 13	*GALNT13*	2q23.3-q24.1	Expressed in the brain, B cells, kidney, and liver as well as potentially being involved in metabolism and energy pathways.	rs10196189	Intron variant A > G	Power athlete status and speed.	G = 0.14
Solute carrier family 16 member 1	*SLC16A1*	1p13.2	A monocarboxylate transporter that catalyses the movement of monocarboxylates, such as lactate and pyruvate.	rs1049434	Missense variant T > A (Asp > Glu)	Lactate transport rate, lactate accumulation, maximum oxygen uptake, power and endurance athlete status.	A = 0.44
Superoxide dismutase 2	*SOD2*	6q25.3	Catalyses the dismutation of superoxide radicals in mitochondria by converting anion superoxide into hydrogen peroxide and oxygen	rs4880	Missense variant A > G (Val > Ala)	MnSOD efficiency against oxidative stress, MnSOD protein formation, CKM activity, power athlete status.	G = 0.47
Thyrotropin releasing hormone receptor	*TRHR*	8q23.1	Stimulates the release of thyrotropin and prolactin, having a role in the regulation of metabolic and hormonal functions.	rs7832552	Intron variant C > T	Lean body mass, strength, and power athlete status.	T = 0.27
Uncoupling protein 2	*UCP2*	11q13.4	Involved in energy expenditure that facilitates the transfer of anions and protons in the inner and outer mitochondrial membrane.	rs660339	Missense variant G > A (Ala > Val)	Energy expenditure, body mass, aerobic and anaerobic capacities, endurance and power athlete status.	A = 0.40

Note. Chr = chromosome location; MAF = minor allele frequency (according to European population; 1000 Genomes Project Consortium, 2015).

9.3 Results

Genotype and allele distributions of all SNPs were in HWE, except for *GALNT13* ($p < .001$) and *UCP2* ($p = .010$) in the PDP group (see **Table 9.2**). The genotype frequency distribution of *IL6* was significantly different between YDP and PDP players ($p = .023$) (see **Figure 9.1**). More specifically, the G allele was overrepresented (13.3%) in PDP players (90.5%) compared to YDP players (77.2%). Furthermore, PDP players had 2.83 times the odds of possessing a G allele (OR = 2.83, 95% CI: 1.13-7.09) compared to YDP players. No significant differences in genotype frequency distribution between the age-specific phases for any other SNPs existed (see **Table 9.3**).

The TGS of players ranged from 31 to 69 in the YDP group and 38 to 73 in the PDP group (see **Figure 9.2**). The mean TGS of PDP players (54.9 ± 8.41) was significantly higher than YDP players (50.6 ± 8.62; $t_{(164)} = 3.26$, $p = .001$). The YDP tertile distribution was: lower = 27, middle = 43, and higher = 22, whereas the PDP tertile distribution was: lower = 16, middle = 24, and higher = 34. Compared to YDP players, PDP players had 2.61 times the odds of having a TGS in the higher third (i.e., 58-73) than a TGS in the lower third (i.e., 31-47; OR = 2.61, CI: 1.15-5.91), as well as 2.77 times the odds of having a TGS in the higher third than a TGS in the middle third (i.e., 48-57; OR = 2.77, CI: 1.33-5.76). The ROC analysis determined that TGS frequency distribution showed significant, but poor, discriminatory power in distinguishing YDP and PDP players (AUC = 0.643, 95% CI: 0.560-0.726).

The TWGS of players ranged from 31 to 74 in the YDP group and 24 to 78 in the PDP group (see **Figure 9.3**). The mean TWGS of PDP players (56.5 ± 9.63) was significantly higher than YDP players (50.0 ± 9.54; $t_{(164)} = 4.34$, $p < .001$). The YDP tertile distribution was: lower = 39, middle = 29, and higher = 24, whereas the PDP tertile distribution was: lower = 11,

middle = 24, and higher = 39. Compared to YDP players, PDP players had 5.76 times the odds of having a TWGS in the higher third (i.e., 57-78) than a TWGS in the lower third (i.e., 24-48; OR = 5.76, CI: 2.49-13.35), as well as 2.93 times the odds of having a TWGS in the middle third (i.e., 49-56) than a TWGS in the lower third (OR = 2.93, CI: 1.24-6.94). The ROC analysis determined that TWGS frequency distribution showed significant, but poor, discriminatory power in distinguishing YDP and PDP players (AUC = 0.694, 95% CI: 0.615-0.773).

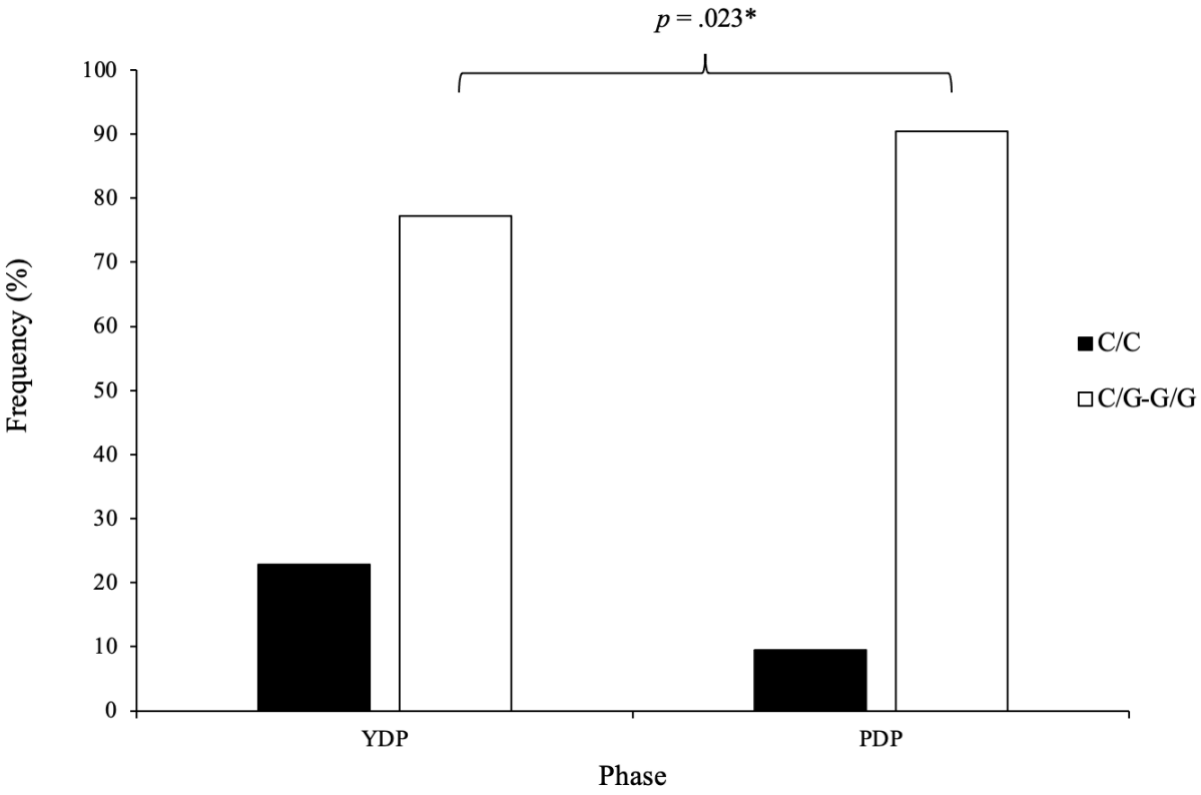


Figure 9.1. The *IL6* (rs1800795) frequency distribution in youth development phase (YDP) and professional development phase (PDP) English academy football players. * Statistically significant at $p < .05$.

Table 9.2. Descriptive statistics of youth and professional development phase English academy football players.

Gene (SNP)	Genotype	YDP = *n* (%)	PDP = *n* (%)	All = *n* (%)	MAF	HWE
HTR2A (rs6311)	C/C	39 (42)	25 (34)	64 (39)	0.40	0.26
	C/T	36 (39)	36 (49)	72 (43)		
	T/T	17 (18)	13 (18)	30 (18)		
ACE (rs4341)	G/G	27 (29)	21 (28)	48 (29)	0.47	0.76
	G/C	45 (49)	35 (47)	80 (48)		
	C/C	20 (22)	18 (24)	38 (23)		
ACTN3 (rs1815739)	C/C	34 (37)	26 (35)	60 (36)	0.39	0.42
	C/T	46 (50)	38 (51)	84 (51)		
	T/T	12 (13)	10 (14)	22 (13)		
ADBR2 (rs1042714)	C/C	27 (29)	20 (27)	47 (28)	0.48	0.44
	C/G	43 (47)	35 (47)	78 (47)		
	G/G	22 (24)	19 (26)	41 (25)		
AGT (rs699)	A/A	26 (28)	24 (32)	50 (30)	0.45	1
	A/G	48 (52)	34 (46)	82 (49)		
	G/G	18 (20)	16 (22)	34 (20)		
AMPD1 (rs17602729)	G/G	74 (80)	59 (80)	133 (80)	0.11	0.70
	G/A	17 (18)	14 (19)	31 (19)		
	A/A	1 (1)	1 (1)	2 (1)		
BDNF (rs6265)	C/C	60 (65)	50 (68)	110 (66)	0.19	1
	C/T	28 (30)	22 (30)	50 (30)		
	T/T	4 (4)	2 (3)	6 (4)		
COMT (rs4680)	G/G	30 (33)	22 (31)	53 (32)	0.43	0.87
	G/A	47 (51)	36 (49)	83 (50)		
	A/A	15 (16)	15 (20)	30 (18)		
CTNNA2 (rs7600563)	T/T	44 (52)	37 (51)	81 (51)	0.28	1
	T/G	34 (40)	31 (42)	65 (41)		
	G/G	7 (8)	5 (7)	12 (8)		
CHRM2 (rs1824024)	A/A	37 (40)	31 (42)	68 (41)	0.36	0.87
	A/C	42 (46)	36 (49)	78 (47)		
	C/C	13 (14)	7 (9)	20 (12)		
	C/C	63 (73)	48 (66)	111 (70)	0.16	0.77

CPNE5 (rs3213537)	C/T	21 (25)	24 (33)	45 (28)		
	T/T	2 (2)	1 (1)	3 (2)		
CKM (rs8111989)	T/T	45 (49)	41 (55)	86 (52)		
	T/C	38 (41)	28 (38)	66 (40)	0.28	0.85
	C/C	9 (10)	5 (7)	14 (8)		
DBH (rs1611115)	C/C	56 (61)	44 (59)	100 (60)		
	C/T	33 (36)	27 (36)	60 (36)	0.22	0.50
	T/T	3 (3)	3 (4)	6 (4)		
DRD1 (rs4532)	T/T	35 (38)	33 (45)	68 (41)		
	T/C	46 (50)	36 (49)	82 (49)	0.34	0.30
	C/C	11 (12)	5 (7)	16 (10)		
DRD2 (rs1076560)	C/C	65 (71)	46 (62)	111 (67)		
	C/A	23 (25)	25 (34)	48 (29)	0.19	0.61
	A/A	4 (4)	3 (4)	7 (4)		
DRD3 (rs6280)	T/T	36 (39)	28 (38)	64 (39)		
	T/C	42 (46)	38 (51)	80 (48)	0.37	0.74
	C/C	14 (15)	8 (11)	22 (13)		
DRD4 (rs1800955)	C/C	20 (25)	14 (22)	34 (23)		
	C/T	32 (40)	32 (49)	64 (44)	0.44	0.18
	T/T	29 (36)	19 (29)	48 (33)		
FTO (rs9939609)	T/T	31 (34)	20 (27)	51 (31)		
	T/A	41 (45)	43 (58)	84 (51)	0.44	0.87
	A/A	20 (22)	11 (15)	31 (19)		
GABRA6 (rs3219151)	T/T	27 (30)	22 (30)	49 (30)		
	T/C	47 (52)	40 (55)	87 (53)	0.44	0.35
	C/C	17 (19)	11 (15)	28 (17)		
GALNT13 (rs10196189)	A/A	63 (68)	46 (62)	109 (66)		
	A/G	23 (25)	18 (24)	41 (24)	0.22	<0.001
	G/G	6 (7)	10 (14)	16 (10)		
HIF1A (rs11549465)	C/C	69 (75)	57 (77)	126 (76)		
	C/T	22 (24)	16 (22)	38 (23)	0.13	1
	T/T	1 (1)	1 (1)	2 (1)		
HSD17B14 (rs7247312)	A/A	72 (78)	62 (84)	134 (81)		
	A/G	17 (18)	12 (16)	29 (17)	0.11	0.39
	G/G	3 (3)	0 (0)	3 (2)		

IGF1 (rs35767)	G/G	65 (71)	44 (59)	109 (66)		
	G/A	26 (28)	27 (36)	53 (32)	0.18	0.60
	A/A	1 (1)	3 (4)	4 (2)		
IGF2 (rs680)	C/C	49 (53)	35 (47)	84 (51)		
	C/T	37 (40)	35 (47)	72 (43)	0.28	0.34
	T/T	6 (7)	4 (6)	10 (6)		
IL6 (rs1800795)	G/G	32 (35)	29 (39)	61 (37)		
	G/C	39 (42)	38 (51)	77 (46)	0.40	0.75
	C/C	21 (23)	7 (9)	28 (17)		
NOS3 (rs2070744)	T/T	37 (40)	28 (38)	65 (39)		
	T/C	42 (46)	40 (54)	82 (49)	0.36	0.41
	C/C	13 (14)	6 (8)	19 (11)		
OXTR (rs2254295)	T/T	68 (79)	57 (78)	125 (79)		
	T/C	15 (17)	14 (19)	29 (18)	0.12	0.06
	C/C	3 (3)	2 (3)	5 (3)		
PPARA (rs4253778)	G/G	55 (60)	52 (70)	107 (65)		
	G/C	33 (36)	17 (23)	50 (30)	0.20	0.34
	C/C	4 (4)	5 (7)	9 (5)		
PPARG (rs1801282)	C/C	77 (84)	60 (81)	137 (83)		
	C/G	15 (16)	12 (16)	27 (16)	0.09	0.64
	G/G	0 (0)	2 (3)	1 (1)		
SLC16A1 (rs1049434)	T/T	32 (35)	23 (31)	55 (33)		
	T/A	46 (50)	37 (50)	83 (50)	0.42	0.87
	A/A	14 (15)	14 (19)	28 (17)		
SOD2 (rs4880)	A/A	26 (28)	19 (26)	45 (27)		
	A/G	40 (43)	40 (54)	80 (48)	0.49	0.64
	G/G	26 (28)	15 (20	41 (25)		
TRHR (rs7832552)	C/C	51 (55)	38 (51)	89 (54)		
	C/T	30 (33)	29 (39)	59 (36)	0.29	0.09
	T/T	11 (12)	7 (9)	18 (11)		
UCP2 (rs660339)	G/G	27 (29)	17 (23)	44 (27)		
	G/A	49 (53)	48 (65)	97 (58)	0.44	0.03
	A/A	16 (17)	9 (12)	25 (15)		

Note. YDP = youth development phase; PDP = professional development phase; MAF = minor allele frequency; HWE = Hardy-Weinberg equilibrium.

Table 9.3. Genetic associations with youth and professional development phase English academy footballers.

Gene (SNP)	Model	YDP (%)	PDP (%)	*B*	OR (95% CI)	*p*
HTR2A (rs6311)	C/C	42	34	1.35	0.69 (0.37-1.31)	.260
	C/T-T/T	58	66			
ACE (rs4341)	G/G	29	28	0.31	0.95 (0.49-1.88)	1
	G/C-C/C	71	72			
ACTN3 (rs1815739)	C/C	37	35	0.28	0.92 (0.49-1.75)	.871
	C/T-T/T	63	65			
ADBR2 (rs1042714)	C/C	29	27	0.24	0.89 (0.45-1.76)	.863
	C/G-G/G	71	73			
AGT (rs699)	A/A	28	32	0.74	1.22 (0.63-2.37)	.611
	A/G-G/G	72	68			
AMPD1 (rs17602729)	G/G	80	80	0.37	0.96 (0.44-2.06)	1
	G/A-A/A	20	20			
BDNF (rs6265)	C/C	65	68	0.11	1.11 (0.58-2.13)	.869
	C/T-T/T	35	32			
COMT (rs4680)	G/G-G/A	84	80	1.57	0.77 (0.35-1.69)	.547
	A/A	16	20			
CTNNA2 (rs7600563)	T/T	52	51	0.42	0.96 (0.51-1.79)	1
	T/G-G/G	48	49			
CHRM2 (rs1824024)	A/A	40	42	0.25	1.07 (0.58-2.00)	.875
	A/C-C/C	60	58			
CPNE5 (rs3213537)	C/C	73	66	0.40	0.70 (0.36-1.38)	.386

	C/T-T/T	27	34			
CKM (rs8111989)	T/T	49	55	0.72	1.30 (0.70-2.40)	.437
	T/C-C/C	51	45			
DBH (rs1611115)	C/C	61	60	0.03	0.94 (0.50-1.76)	.874
	C/T-T/T	39	40			
DRD1 (rs4532)	T/T	38	45	0.02	1.31 (0.70-2.44)	.430
	T/C-C/C	62	55			
DRD2 (rs1076560)	C/C	71	62	1.26	0.68 (0.36-1.31)	.320
	C/A-A/A	29	38			
DRD3 (rs6280)	T/T	39	38	0.32	0.95 (0.50-1.78)	.874
	T/C-C/C	61	62			
DRD4 (rs1800955)	C/C	25	22	0.36	0.84 (0.38-1.82)	.697
	C/T-T/T	75	78			
FTO (rs9939609)	T/T	34	27	0.93	0.73 (0.37-1.43)	.400
	T/A-A/A	66	73			
GABRA6 (rs3219151)	T/T-T/C	81	85	0.33	1.29 (0.56-2.97)	.677
	C/C	19	15			
GALNT13 (rs10196189)	A/A	68	62	0.01	0.76 (0.40-1.44)	.415
	A/G-G/G	32	38			
HIF1A (rs11549465)	C/C	75	77	0.13	1.12 (0.54-2.29)	.856
	C/T-T/T	25	23			
HSD17B14 (rs7247312)	A/A	78	84	0.63	1.44 (0.65-3.17)	.431
	A/G-G/G	22	16			
IGF1 (rs35767)	G/G	71	59	1.34	0.61 (0.32-1.16)	.142
	G/A-A/A	29	41			

IGF2 (rs680)	C/C	53	47	0.94	0.79 (0.43-1.45)	.532
	C/T-T/T	47	53			
IL6 (rs1800795)	G/G-G/C	77	91	3.00	2.83 (1.13-7.09)	**.023***
	C/C	23	9			
NOS3 (rs2070744)	T/T	40	38	0.21	0.90 (0.48-1.70)	.873
	T/C-C/C	60	62			
OXTR (rs2254295)	T/T	79	78	0.07	0.94 (0.44-2.02)	1
	T/C-C/C	21	22			
PPARA (rs4253778)	G/G	60	70	0.00	1.59 (0.83-3.05)	.193
	G/C-C/C	40	30			
PPARG (rs1801282)	C/C	84	81	1.49	0.83 (0.37-1.86)	.685
	C/G-G/G	16	19			
SLC16A1 (rs1049434)	T/T	35	31	0.68	0.85 (0.44-1.62)	.624
	T/A-A/A	65	69			
SOD2 (rs4880)	A/A-A/G	72	80	0.64	1.55 (0.75-3.20)	.279
	G/G	28	20			
TRHR (rs7832552)	C/C	55	51	0.11	0.85 (0.46-1.57)	.640
	C/T-T/T	45	49			
UCP2 (rs660339)	G/G	29	23	0.48	0.72 (0.36-1.45)	.381
	G/A-A/A	71	77			

Note. Bold values and * highlight statistical significance at $p < .05$. YDP = youth development phase; PDP = professional development phase; *B* = unstandardised beta; OR = odds ratio; CI = confidence interval.

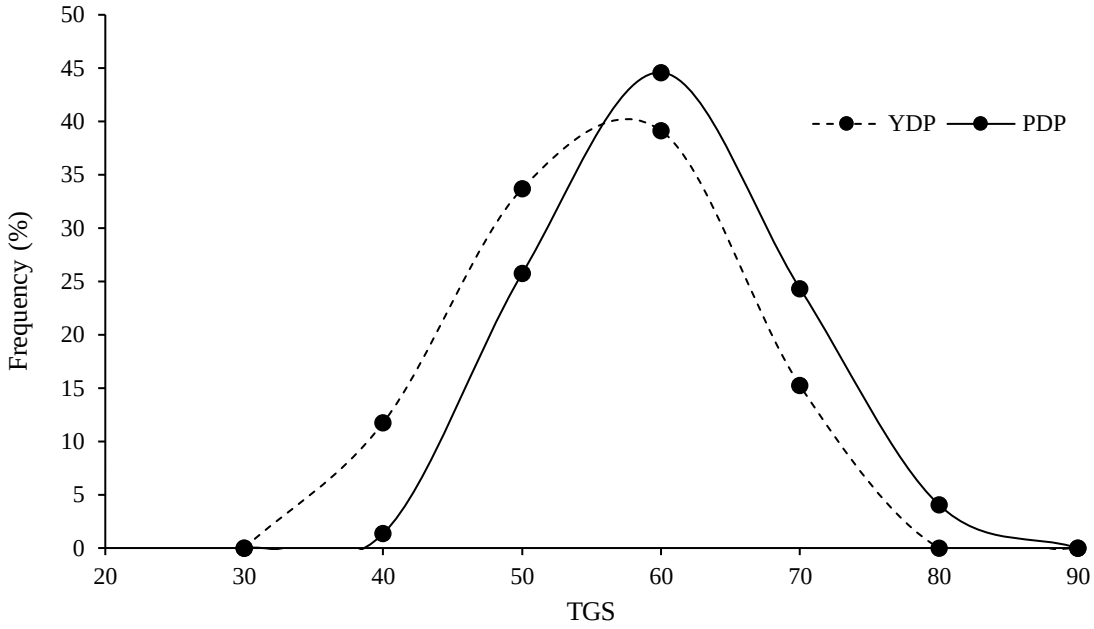


Figure 9.2. Frequency distribution of total genotype score (TGS) in youth development phase (YDP) and professional development phase (PDP) English academy football players.

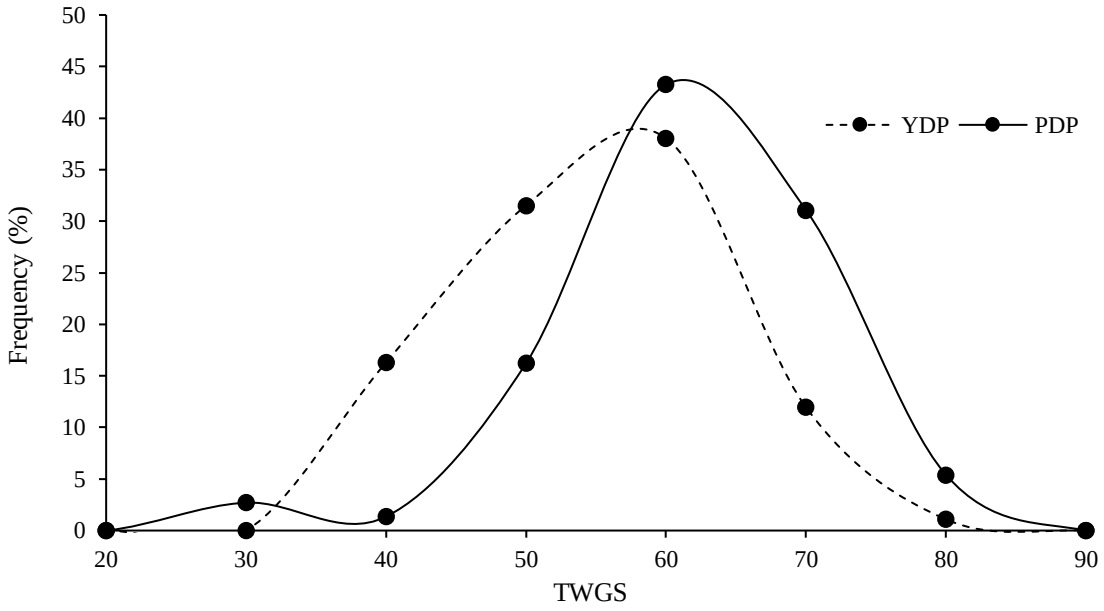


Figure 9.3. Frequency distribution of total weighted genotype score (TWGS) in youth development phase (YDP) and professional development phase (PDP) English academy football players.

9.4 Discussion

This study examined differences in the genotype frequency distribution of thirty-three SNPs, both individually and collectively, between YDP and PDP English academy football players. Key findings showed an overrepresentation of the *IL6* (rs1800795) G allele in PDP players compared to YDP players. In addition, the T(W)GS models demonstrated that the combination of these thirty-three SNPs was effective in differentiating YDP and PDP players. As such, these results suggest there is significant genetic variation between youth football players of distinct age groups. To the author's knowledge, this is the first assessment of genotype frequency distribution in isolation, and as part of a polygenic profile, between two age-specific phases of academy football players in England. Therefore, these findings may have important implications for future studies examining genetic associations in youth football.

The *IL6* gene encodes for the pleiotropic cytokine interleukin-6 (IL-6), previously associated with multiple biological processes relevant to sport performance (i.e., glucose homeostasis, muscle hypertrophy, and repairing damaged muscle) (Yamin et al., 2008). The circulating levels of IL-6 can vary depending on specific variants within the gene. For instance, the G and C alleles of the *IL6* (rs1800795) SNP alter promoter activity and consequently result in higher and lower IL-6 levels, respectively (Fishman et al., 1998). Higher IL-6 levels have been associated with greater muscle hypertrophy, improved glucose uptake, and increased protection against exercise-induced muscle damage, possibly due to reduced muscle inflammation by positively regulating the pro- and anti-inflammatory cytokine production balance (Carey et al., 2006; Petersen & Pedersen, 2005). In contrast, lower IL6 levels may increase the possibility of sustaining a muscular injury, inhibit recovery, and hinder athletic

performance, with higher creatine kinase activity reported in response to eccentric exercise in C allele carriers (Yamin et al., 2008).

More recent sport-specific research has shown that IL-6 may be an important biomarker in power-orientated sports and performance phenotypes. Studies assessing Polish and Spanish high performing athletes have reported an overrepresentation of the *IL6* (rs1800795) G allele in those who take part in power-based sports (i.e., jumpers, sprinters, and weightlifters) compared to controls (Eider et al., 2013; Ruiz, Buxens, et al., 2010). Cross-sectional quantitative data supporting these findings also exist, as youth footballers in Britain possessing the G allele performed significantly better than C allele carriers in acceleration and speed assessments (i.e., 5 m and 20 m sprint) (Pickering et al., 2019). Therefore, due to the mechanistic properties associated with *IL6* (rs1800795), the G allele may better protect skeletal muscle and aid in repair during powerful muscle contractions, which subsequently allows for a higher volume of training that stimulates favourable adaptations, and ultimately results in superior performance in high-intensity activities.

Although power-orientated phenotypes such as acceleration, speed, and vertical jumps are important across all youth football age groups (Kelly & Williams, 2020), they appear to become *more* important as players age and mature (Murtagh et al., 2018). For instance, in many male English football academies, youth players do not progress to compete on a full-sized pitch, with eleven players on each team, until the under-13 age group. With this increase in pitch size, players spend more of their competitive match-play time at low speeds and perform a greater number of sprint actions, placing a greater physiological demand on anaerobic capacity (Casamichana et al., 2012). Furthermore, in a longitudinal investigation of English academy football players, it was reported that whilst future professionals began to outperform

their non-professional counterparts in vertical CMJ from the age of 12 years (> 0.6 cm), differences became more pronounced in older age groups (e.g., aged 18 years > 1.7 cm) (Saward et al., 2020).

As competitive match-play demands shift more towards anaerobic capacities, academy recruitment teams may choose to retain players displaying superior power rather than endurance capabilities (Murtagh et al., 2018). This may explain the overrepresentation of the *IL6* (rs1800795) G allele in the PDP group compared to the YDP group due to its association with several power-orientated phenotypes. However, recent research in academy football has also shown that the G allele may protect PDP players from injury. More specifically, Hall et al. (2022) reported that only post-peak height velocity players (aged 17.5 ± 2.1 years) possessing the *IL6* (rs1800795) C/C genotype suffered significantly more injuries than G allele carriers. The authors noted that the association was possibly due to the combination of greater muscle damage and inflammation experienced by C allele carriers, alongside the higher-intensity of match actions, and increased frequency of training and/or competitive match-play in older age groups. As such, the overrepresentation of the G allele in the PDP group may be explained by a pleiotropic effect of *IL6* (rs1800795) on power and injury.

The TGS and TWGS models showed that YDP and PDP football players have distinct polygenic profiles, with the TWGS demonstrating greater discriminatory accuracy. This suggests that whilst each SNP has a small additive effect, favourable alleles of individual SNPs have different degrees of influence. This corresponds with previous research in academy football players on physiological, psychological, and technical phenotypes that underpin differences in these age-specific phases (see McAuley, Hughes, et al., 2022a, 2022b; Murtagh et al., 2020). The general frequency distribution of the genotypes across all SNPs also aligns

with the *IL6* (rs1800795) findings. Specifically, PDP players had a greater proportion of alleles previously associated with power-orientated phenotypes. This indicates PDP players have an overall more power-orientated polygenic profile, which corresponds with similar findings reported in post-peak height velocity (aged 16.8 ± 2.3 years) academy football players using only four of these SNPs: *ACTN3* (rs1815739), *AGT* (rs699), *PPARA* (rs4253778), and *NOS3* (rs2070744) (Murtagh et al., 2020).

The polygenic models also showed that in general YDP players had a greater proportion of favourable alleles in SNPs previously associated with psychological (i.e., personality dimensions and mental toughness) and technical (i.e., dribbling, passing, and shooting) phenotypes in academy footballers (see McAuley et al., 2022a, 2022b). The importance of these psychological and technical phenotypes in youth football has been demonstrated in previous research by effectively differentiating higher and lower performers in adolescence and predicting success at adulthood (Sarmento et al., 2018; Williams et al., 2020). However, these findings suggest having an increased frequency of these preferred psychological/technical alleles may be more advantageous in younger age groups. This corresponds with previous research that reported coaches and recruiters consider technical, tactical, and psychological factors as the most important during this stage of development (Fuhre et al., 2022; Larkin & O'Connor, 2017). As such, the polygenic models collectively showcase that English academy football players of different age-specific phases may have distinct genetic profiles, with PDP players more power-orientated and YDP players more psychological- and technical-orientated.

Although the polygenic models distinguished YDP and PDP players, they still had relatively poor accuracy, which indicates they should not be considered for practical

implementation. Moreover, given the data-driven cross-sectional nature of the analyses, these findings may not generalise well to other youth football cohorts and may reflect cohort effects. Therefore, the external validity of these results should be assessed in larger independent samples alongside the addition of many more relevant genetic variants. It is also important to note that the previous associations of the SNPs included in this study with specific physiological, psychological, technical, and injury phenotypes may not be reliable due to the relatively small sample sizes in football genomic research (McAuley, Hughes, et al., 2021a). Therefore, the inferences made with regards to genetic profile orientation in YDP and PDP players should be interpreted with caution.

Studies with this type of unique sample are typically underpowered so it is important to be relatively conservative with any conclusions, as meaningful implications cannot be made from one study in isolation. However, in the early stages of development in a field, informed speculation based on prior knowledge may be important for informing future work. As a result, we made informed speculation about our findings as a way of guiding subsequent work in this area. Moreover, building this research base with studies using transparent methodologies is important so they can contribute to research synthesis approaches in the future, and draw more valid and reliable conclusions before these findings are implemented into applied settings (McAuley, Baker, et al., 2022).

9.5 Conclusion

This study has presented novel evidence with regards to the genetic profiles of YDP and PDP male academy football players in England. To be specific, the *IL6* (rs1800795) G allele was overrepresented in PDP players compared to YDP players, possibly due to its theorised pleiotropic effect on power and injury phenotypes. Moreover, the T(W)GS models derived

from all thirty-three SNPs effectively distinguished YDP and PDP players, with PDP players exhibiting an overall more power-orientated polygenic profile. As such, this study has shown for the first time that there is significant inter-individual genetic variation between youth football players of specific age phases in English academies. If validated in larger independent youth football cohorts, these findings may have important implications for future studies examining genetic associations in youth football.

10 Integrated discussion

Considerable evidence now exists that suggests there is a significant genetic component to the various parameters that underpin athlete development and performance, even when adjusting for environmental effects (Georgiades et al., 2017; Yan et al., 2016). Indeed, heritability studies have reported that a range of anthropometric (e.g., height) physiological (e.g., power), psychological (e.g., personality), technical (e.g., motor control), and injury traits (e.g., ACL rupture), as well as overall athlete status are moderately to highly hereditary (De Moor et al., 2007; Horsburgh et al., 2009; Magnusson et al., 2020; Missitzi et al., 2013; Polderman et al., 2015; Silventoinen et al., 2008; Zempo et al., 2017). Whilst heritability studies establish a foundation of the potential genetic influence on human traits, they do not reveal which specific biological variants directly contribute to the observed inter-individual differences (Georgiades et al., 2017; Guilherme et al., 2014). However, more sophisticated molecular biology techniques developed over the past two decades have enabled the analysis of specific genetic variants and their association with selected phenotypes (Pitsiladis & Wang, 2011).

The focus of current genetic research is on further understanding these genotype-phenotype relationships using various genetic association approaches (e.g., CGAS, GWAS, and TGS) (Guilherme et al., 2014; Visscher et al., 2012; Williams & Folland, 2008). However, the majority of genetic association research has centred on individual sports (Ahmetov et al., 2016), which means there is currently a lack of studies focused on associations in team sports such as football. Since sports vary greatly in their defining characteristics and contextual demands, it is likely there are also significant distinctions at the molecular level between athletes of different sports (Guilherme et al., 2014; Heffernan et al., 2015). As such, genetic associations cannot be generalised across all sport settings and need to be investigated with

independent homogenous cohorts (Mattsson et al., 2016). Accordingly, the overarching aim of this thesis was to investigate the association of genetic polymorphisms with phenotypes in football-specific contexts to enhance understanding of the underpinning biological mechanisms, which may contribute towards facilitating greater individualised athlete development in the future.

The purpose of this chapter is to provide an integrative discussion, whereby the key points of the individual articles are summarised and amalgamated to offer a cohesive interpretation of this project's findings and highlight novel contributions to the research field. Limitations of the project, proposed directions for future research, and implications for practice are also outlined before concluding.

10.1 Research syntheses section

The purpose of the first study (Chapter 2) was to identify and synthesise genetic association studies involving football players to assess the current progress and methodological rigor of this research base by conducting a systematic review of the published literature. This provided a comprehensive scoping evaluation of the present methodological protocols and study designs adopted by researchers, as well as the genes and phenotypes investigated within the discipline. Moreover, this facilitated the identification of major limitations in this research field that need to be addressed and outlined future research avenues that could be explored in subsequent chapters. To the author's knowledge, this is the first review of genetic association studies specific to a football context, providing a timely insight into an expanding area of research and a novel contribution to the scientific literature.

In summary, 80 eligible studies were identified, of which 55% (n = 44) were published within the last four years at the time of the search (i.e., September 2019), showcasing that

genetic association research in football has increased at a substantial rate. Studies predominately focused on the association of the *ACTN3* or *ACE* gene with athlete status, used case-control CGAS designs with senior male cohorts, grouped athletes of different sports together with footballers, used participants of different defining characteristics (e.g., competitive playing level, ethnicity, sex, and on-field position), and had a median sample size of 60. Of the sub-disciplines that were explored, some already required an independent review (i.e., injury), whilst others required supplementary empirical research (i.e., psychological). Furthermore, there were several novel fields of research yet to be investigated in footballers (i.e., technical skill).

Overall, this review identified several methodological limitations currently preventing conclusive evaluations (i.e., cohort heterogeneity, population stratification, and sample size). In some cases there were also inadequate descriptions or complete omissions of information pertaining to participant characteristics. Therefore, adherence to STREGA guidelines (Little et al., 2009), transparent and consistent reporting, and a greater awareness of cohort individualities are advised. In addition, due to the relatively small sample sizes and the predominant use of case-control CGASs, larger sample sizes, cross-sectional and/or longitudinal study designs, and the utilisation of genome-wide and/or polygenic profiling approaches are recommended. It is important to note the difficulty associated with some of these limitations (i.e., sample size, ethnic conformation, and GWAS). Therefore, researchers should view these recommendations as a guide towards future best practice, when more favourable conditions are available.

The purpose of the second study (Chapter 3) was to assess the association of the *ACTN3* R577X and *ACE* I/D polymorphisms with athlete status in football via a meta-analysis.

Although the first study highlighted the association of *ACTN3* and *ACE* with athlete status is the most studied within football genomics, there was no general consensus on their importance as studies reported contrasting allelic associations (e.g., Egorova et al., 2014; Galeandro et al., 2017; Santiago et al., 2008; Ulucan et al., 2015). This is most likely because each allele has a small contribution to performance alongside studies having small heterogenic cohorts (Monnerat et al., 2019; Wang et al., 2013). Therefore, to overcome these limitations, a meta-analysis was used to pool the results of small homogenous studies together. To the author's knowledge, this is the first meta-analysis of genetic associations with athlete status in a football-specific context. Moreover, it is the first meta-analysis of a homogenous team-sport cohort, representing a significant original contribution to this field of research.

To summarise, the systematic search identified 17 *ACTN3* and 19 *ACE* studies that presented the genotype frequencies of footballers in isolation. The *ACTN3* studies comprised 1759 football players, whereas the *ACE* studies comprised 1925 football players. Significant associations were shown between presence of the *ACTN3* R allele and professional footballer status and the *ACE* D allele and youth footballers, compared to a control group. These findings may be explained by the combination of several factors: (a) the number and frequency of powerful actions performed in a game (i.e., 1000-1400 acyclical bursts of activity, including; jumps, tackles, shots at goal, changes of direction, and, sprints) (Bradley et al., 2009; Stølen et al., 2005), (b) the contribution of the ability to repeat higher intensity actions to success in football (i.e., league position, goals scored, goals prevented, duel success) (Little & Williams, 2005; Oliver et al., 2009), and (c) the association of the *ACTN3* R allele and *ACE* D allele with power-orientated phenotypes (i.e., acceleration, speed, and vertical jump height) (Dionísio et al., 2017; Pimenta et al., 2013).

In conclusion, the results of this meta-analysis provide further evidence that inter-individual genetic variation likely contributes towards athlete status in football. Moreover, the findings suggest that genetic variation can differentiate athletes of different competitive playing statuses in a homogenous team-sport cohort. Specifically, this study has showed in the largest sample of footballers to date, the *ACTN3* R577X and *ACE* I/D polymorphisms are likely (albeit relatively minor) contributing factors which influence athlete status in football. Cross-sectional and longitudinal studies are required to establish a strong evidence base regarding the specific phenotypes influenced by these polymorphisms and how different interventions can optimise development based on genotype. Once a strong evidence base has been established which supports genetic-based programme design, these results could be used alongside a number of other extensively evidenced/researched genotypes to form the basis of genetic-based training.

The purpose of the third study (Chapter 4) was to synthesise genetic association studies investigating injury involving football players to identify which genetic variants have the most empirical evidence. The first study revealed that injury was the most examined phenotype in football genomic research, however it was beyond the scope of that review to perform an in-depth analysis of the allelic associations within the context of the intra- and inter-study limitations and variability. Identifying risk factors that predispose injury is important, not only for player welfare and performance, but also because of the profound effect on a football team's potential success and economic burden (Eirale et al., 2013; Eliakim et al., 2020; Hägglund et al., 2013). To the author's knowledge, whilst injury-specific genetic association reviews have been completed in other team-sports such as rugby (Brazier et al., 2019), no study reviewed genetic associations with injury specifically in football players. As such, this study represents a novel addition to this research field and a meaningful advancement in knowledge.

Due to the observed heterogeneity between studies, a statistically pooled quantitative synthesis of the extracted data could not be performed so a narrative synthesis summarised results. Firstly, the systematic search process culminated in the inclusion of 34 studies, of which there were 33 CGAS and one GWAS. Across the 33 CGASs, a total of 99 unique polymorphisms were assessed within 63 genes. Forty-one unique polymorphisms were associated with injury at least once, whereas three polymorphisms had their specific allelic associations with injury replicated at least twice in independent cohorts. More specifically, *ACTN3* (rs1815739) X/X genotype was associated with an increased susceptibility to non-contact muscle injuries, *ACAN* (rs1516797) G allele was associated with increased susceptibility to ACL injuries, and *VEGFA* (rs2010963) C/C genotype was associated with an increased susceptibility to ACL and ligament or tendon injuries.

However, when critically analysing the methodological approach and cohort characteristics of each *ACTN3*, *ACAN*, and *VEGFA* study, there were some significant within- and between-study limitations and variability limiting the subsequent reliability and external validity of these findings. Moreover, some of the issues present in these studies (e.g., small sample sizes, cohort heterogeneity, and population stratification) were prevalent throughout all studies identified in this review. As such, within a football injury context, currently there doesn't appear to be any replicated and validated genetic variants. Therefore, although the utilisation of genetic information to manage injury susceptibility in football is an exciting prospect, the evidence base supporting the use of genetic testing as a prognostic or diagnosis tool is weak. Future research via collaborations with improved methodological approaches and mechanistic studies may help identify significant biological pathways underpinning injury risk.

Although, to truly have meaningful clinical application, genetic information will have to be used collectively alongside the various other intrinsic and extrinsic risk factors of injury.

10.2 Experimental section

The purpose of the first experimental study (Chapter 5) was to assess the current practical application of genetic testing in professional football and provide an insight into the perspectives of key stakeholders (i.e., coaches, practitioners, players). The initial scoping review of the literature identified that there was little evaluation regarding the extent of genetic testing in football. More specifically, there were only two peer-reviewed studies that assessed the current use and opinions of genetic testing in sport (Pickering & Kiely, 2021; Varley et al., 2018). However, these studies only included 23 and 22 stakeholders employed in football respectively, limiting the application of their results to a football-specific context. To the authors' knowledge, this is the first study to quantify the use and opinions of genetic testing from a wide range of key stakeholders specifically in football. Moreover, it is the first survey in sport genomics to gather qualitative data alongside quantitative data to facilitate a deeper understanding of the reasons behind responses, providing a significant original contribution to the scientific literature.

In summary, 122 participants completed an online anonymous survey. This consisted of 21 multiple choice and Likert scale questions, with the option of providing an explanation for each response. Findings revealed genetic testing is rarely utilised by key stakeholders (10%) or their respective organisations (14%). However, three quarters (75%) had the opinion that genetic testing will have great utility in the future. The majority (72%) believed genetic testing should be used for athlete development and injury risk, whilst 35% believed that genetic testing should be utilised for talent identification purposes. However, most key stakeholders viewed

their own (89%) and their colleagues' (79%) knowledge related to genetic testing as insufficient; mainly due to ineffective current communication methods (91%). Most believed educational workshops are required (71%), whilst nearly all (91%) were interested in developing their expertise on the utility of genetic testing.

In conclusion, whilst this study suggests that genetic testing is rarely used within professional football, key stakeholders anticipate that genetic testing will be utilised more in the future. Given the perceived lack of knowledge and education, implementation of education programmes may prove valuable in improving key stakeholders' knowledge and the practical application of genetic testing in professional football. Especially since over a third of participants believed genetic testing should be utilised for talent identification and selection, even though using genetic testing for such purposes is considered immoral, unethical, and may lead to legal issues associated with genetic discrimination (Patel & Varley, 2019; Vlahovich et al., 2017; Webborn et al., 2015). Future studies should design and implement methods of education, and evaluate which are the most effective in providing key stakeholders in football with evidenced-based information on genetic research and testing.

The purpose of the second, third, fourth, and fifth experimental studies (Chapters 6 - 9) was to examine the association of several polymorphisms, with technical, psychological, physiological, and age phase phenotypes in English academy male football players, respectively. The first review found there was a lack of cross-sectional studies with youth cohorts investigating quantitative traits and development. Discovering factors that underpin technical abilities, psychological characteristics, and physiological capacities are important due to their association with the relative success of a team, and their discriminative and prognostic power in youth football (Sarmento et al., 2018; Williams et al., 2020). Moreover, the

importance of these performance phenotypes during development in football appears to alter depending on age (Murtagh et al., 2018; Saward et al. 2020), which means the genetic profiles of youth players may also vary between different developmental phases.

Findings revealed *ADBR2* (rs1042714), *BDNF* (rs6265), *DBH* (rs1611115), and *DRD1* (rs4532) were associated with objective dribbling and shooting assessments. In addition, *DRD3* (rs167771) was associated with levels of agreeableness, whereas *GALNT13* (rs10196189) was associated with 5 m, 10 m, and 20 m sprints. Moreover, *IL6* (rs1800795) was associated with countermovement jump and age phase. The data-driven polygenic profiles were also associated with football phenotypes in all studies, but the TWGS consistently displayed stronger associations than the TGS. This suggests that whilst each polymorphism has a small additive effect on football phenotypes, advantageous alleles of a given polymorphism have different degrees of influence. Overall, these results suggest genetic polymorphisms explain a small proportion of the inter-individual variance in quantitative football phenotypes, and there is significant genetic variation between youth football players of distinct age groups.

To the author's knowledge, these studies represent the first investigations of genetic associations with technical, psychological, and age phase phenotypes within a football-specific context. Moreover, the physiological and age phase studies used the most comprehensive polygenic profiles (i.e., comprised the largest number of polymorphisms) within football genomics to date. Therefore, these studies greatly expand knowledge in this research field and provide novel contributions to the scientific literature. If validated in larger independent youth football cohorts, these findings may aid practitioners in the future when individualising training programs to optimise athlete development. However, the identification of more genetic variants

relevant to these and other football phenotypes is also necessary to improve the sensitivity and specificity of the polygenic profiles.

10.3 Limitations

The specific limitations of the separate studies within this thesis have already been provided at the individual article level. In summary, the research synthesis studies were primarily limited by database and language restrictions and the inclusion of only peer-reviewed publications. Future reviews should search more databases, include studies in languages other than English, and possibly incorporate grey literature. The first study of the experimental section did amass the opinions from a wide-range of coaches, practitioners, and players, however the sample was still relatively small, predominately male, and the questions in the survey were not validated. Moving forward, future research should consider the use of a validated survey, try to recruit larger samples, and capture the views of more female stakeholders. The other four experimental studies shared similar limitations. For instance, the sample sizes were relatively small and some confounding variables could not be controlled for (e.g., ethnicity, maturation status). As high-performance football cohorts are notoriously difficult to gain access to, and are generally small and heterogenic by nature, researchers in the future should consider joining an international football consortium to facilitate improved methodological approaches.

10.4 Future directions

Aside from addressing the limitations noted above, researchers in the future should also explore some of the areas that were beyond the scope of the current thesis. For instance, although a multidisciplinary approach was adopted, genetic associations with the tactical dimension (i.e., perceptual-cognitive expertise) of football performance could not be examined. Furthermore, the initial scoping assessment of the literature found that only three of the 80 identified studies

in this research field have been dedicated to female football players. Moreover, only five studies analysed players according to their on-field position. Female football players were not able to be recruited for this project and on-field positions were not considered due to youth players occupying several playing positions before settling on one at older ages. If sample sizes are able to be dramatically increased, the incorporation of more advanced genomic approaches such as GWAS and whole-genome sequencing to facilitate the discovery of novel polymorphisms and rare variants will also be important. Moreover, further research is required on epigenetic modifications (e.g., DNA methylation) and epistatic interactions.

10.5 Practical implications

The prospective utility of genetic information in applied settings within football appears to have great potential. However, the current scientific evidence base does not yet support the implementation of genetic tests to enhance overall athlete development processes. The findings of this thesis have widened research avenues by showcasing the possible association of specific genetic variants and polygenic profiles with phenotypes underpinning football performance and development. Although, the results of these studies have to be replicated in larger independent cohorts and many more relevant genetic variants must be identified before practical applications can be contemplated. Once a strong evidence base exists with regards to specific genotype-phenotype relationships using cross-sectional study deigns, longitudinal studies will be required to establish how different environmental interventions can optimise development based on genotype. It will be important to enhance the knowledge of key stakeholders concerning the nuances of genetic information and progress in this emerging field by implementing education programs within alacritous organisations.

10.6 Reflections and prospects

In the future, it is likely additional genetic variants will be discovered and incorporated into polygenic profiling tools with more accurate weightings to enhance prognostic capabilities (Pickering et al., 2019). As these significant genomic advances are made, genetic information may have more utility in sport, but not necessarily for talent identification. All human traits are thought to be influenced by genetic factors, although there is considerable between-trait-variance in heritability estimates (Polderman et al., 2015). Moreover, nearly all traits are complex (i.e., influenced by more than one gene) and even those with high heritability are still influenced by a very large number of genetic variants (e.g., height; Yengo et al., 2018).

As discussed, athletic performance is a complex trait underpinned by a multitude of other complex traits, each regulated by an intricate network of interconnected biological pathways and independently influenced by environmental exposures (McAuley, Baker, et al., 2021; Tucker & Collins, 2012). In addition, the degeneracy of biological systems (i.e., structurally different elements performing similar functions) and compensatory nature of athlete development (i.e., weaknesses in some attributes compensated by strengths in others), coupled with the characteristics that encompass optimal performance changing over time (Davids & Baker, 2007; Baker et al., 2019), means distinct genetic profiles will likely be associated with expertise in the sporting domain. Moreover, will these genetic profiles ever be predictive a-priori based on evidence-driven analyses and not just a-posteriori as a result of data-driven analyses? This is before considering simple serendipity (i.e., low probability events). Overall, athletic performance is a qualitative phenotype, not a physiological measurement (Mattsson et al., 2016).

Practitioners are encouraged to utilise research that has established genetic associations with specific physiological, psychological, and injury traits to improve athlete development (e.g., modify training interventions to elicit optimal adaptations, manage athlete welfare, and reduce injury susceptibility). However, practitioners should always remember that genetic information should not be used in an isolated deterministic manner, but rather as an additional objective tool that may help with subjective development decisions; similar to how anthropometric measures, fitness tests, performance analysis statistics, and psychological profiles are currently used (McAuley, Baker, et al., 2021). It has also been proposed that genetic testing may aid in screening young athletes for risk of cardiomyopathies (e.g., hypertrophic cardiomyopathy) and channelopathies (e.g., congenital long-QT syndrome), which are major causes of sudden cardiac death (Tanisawa et al., 2020). Genetic information itself should not be feared, however, fears regarding deterministic misunderstandings and reductionist applications are warranted.

Even if the complexity of processes such as epistasis are fully untangled, environmental contingencies can always alter phenotype outcomes as no trait is 100% heritable (Polderman et al., 2015). Genetic information is simply one piece of the developmental puzzle within a holistic bioecological approach (i.e., an additional column in an athlete profile spreadsheet). There are still appropriate concerns with regards to the area of the genome screened in genetic tests, as some genetic variants could reveal serious health conditions, not only of the athlete but of immediate family members (e.g., cystic cbrosis, Huntington's disease, sickle cell anaemia; Tanisawa et al., 2020). Moreover, screening unnecessary genomic regions increases the possibility of reidentifying anonymised genotype data. Therefore, it is imperative that only

variants with explicit relevance to the chosen sporting trait (which are still yet to be determined) are included in a genetic panel.

Moving forward it will be important to try to grow the Football Gene Project into an established national consortium possibly by networking with other researchers and organisations that have genotyped youth footballers in the UK (e.g., Hall et al., 2022; Murtagh et al., 2020; Pickering et al., 2019). In addition, with the anticipated rise in the utilisation of genetic testing, coupled by the lack of knowledge by stakeholders in football, future research related this project should investigate the value of educational resources as well as more accessible and/or digestible summaries of the genetic literature. The misunderstanding of genetic information is of concern as it could have a profound impact on important sporting choices. Indeed, deterministic views may hold negative consequences as beliefs about the inherent origin of talent affect individuals' behaviour, motivation, and performance. For athletes, this may reflect their willingness to train, level of effort, and response to failure, whereas in coaches, this may reflect their attentiveness, patience, and time dedicated to the development of individual athletes (Baker et al., 2018). How stakeholders would respond after being informed of their own or others genetic information is another important research avenue that should be explored.

10.7 Conclusion

The overarching aim of this thesis was to investigate the association of genetic polymorphisms with phenotypes in football-specific contexts. In doing so, this thesis has provided a holistic and novel addition to the scientific literature through the completion of the multidisciplinary studies contained within. The results of these studies suggest inter-individual genetic variation does influence football phenotypes and has identified several novel associations that warrant

further investigation in larger independent football cohorts. If replicated and validated, the data generated by this thesis may prove to be a useful starting point in establishing a genetic profile tool in the future capable of assisting practitioners with implementing individualised athlete development programs. However, genetic information should not be seen as an isolated determinant, but rather as an additional objective tool with which to enhance subjective development decisions. Genetic information is simply one piece of the developmental puzzle within a holistic bioecological approach (i.e., an additional column in the athlete profile spreadsheet). The formation of an international football consortium will be important to enable the sharing of resources between researchers and facilitate superior methodological approaches in future genetic association research in football. Furthermore, key stakeholders in football are recommended to act with caution if utilising genetic testing, whilst researchers are encouraged to design, implement, and evaluate methods of educational support.

11 References

1000 Genomes Project Consortium, Auton, A., Brooks, L. D., Durbin, R. M., Garrison, E. P., Kang, H. M., Korbel, J. O., Marchini, J. L., McCarthy, S., McVean, G. A., & Abecasis, G. R. (2015). A global reference for human genetic variation. *Nature*, *526*(7571), 68–74. https://doi.org/10.1038/nature15393

Abarghoueinejad, M., Baxter-Jones, A. D. G., Gomes, T. N., Barreira, D., & Maia, J. (2021). Motor performance in male youth soccer players: A systematic review of longitudinal studies. *Sports*, *9*(4), 53. https://doi.org/10.3390/sports9040053

Ahmetov, I., Hall, E., Semenova, E., Pranckeviciene, E., & Valentina, G. (2021). Advances in sports genomics. *Advances in Clinical Chemistry*. https://doi.org/10.1016/bs.acc.2021.07.004

Ahmetov, I. I., Druzhevskaya, A. M., Lyubaeva, E. V., Popov, D. V., Vinogradova, O. L., & Williams, A. G. (2011). The dependence of preferred competitive racing distance on muscle fibre type composition and ACTN3 genotype in speed skaters. *Experimental Physiology*, *96*(12), 1302–1310. https://doi.org/10.1113/expphysiol.2011.060293

Ahmetov, I. I., Egorova, E. S., Gabdrakhmanova, L. J., & Fedotovskaya, O. N. (2016). Genes and athletic performance: An update. *Medicine and Sport Science*, *61*, 41–54. https://doi.org/10.1159/000445240

Ahmetov, I. I., Egorova, E. S., & Mustafina, L. J. (2013). The PPARA gene polymorphism in team sports athletes. *Central European Journal of Sport Sciences and Medicine*, *1*(1), 19–24.

Ahmetov, I. I., & Fedotovskaya, O. N. (2015). Current progress in sports genomics. *Advances in Clinical Chemistry*, *70*, 247–314. https://doi.org/10.1016/bs.acc.2015.03.003

Akaneya, Y., Tsumoto, T., Kinoshita, S., & Hatanaka, H. (1997). Brain-derived neurotrophic factor enhances long-term potentiation in rat visual cortex. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *17*(17), 6707–6716.

Alfieri, A., Martone, D., Randers, M. B., Labruna, G., Mancini, A., Nielsen, J. J., Bangsbo, J., Krustrup, P., & Buono, P. (2015). Effects of long-term football training on the expression profile of genes involved in muscle oxidative metabolism. *Molecular and Cellular Probes*, *29*(1), 43–47. https://doi.org/10.1016/j.mcp.2014.11.003

Ali, A. (2011). Measuring soccer skill performance: A review. *Scandinavian Journal of Medicine & Science in Sports*, *21*(2), 170–183. https://doi.org/10.1111/j.1600-0838.2010.01256.x

Al-Khelaifi, F., Diboun, I., Donati, F., Botrè, F., Abraham, D., Hingorani, A., Albagha, O., Georgakopoulos, C., Suhre, K., Yousri, N. A., & Elrayess, M. A. (2019). Metabolic GWAS of elite athletes reveals novel genetically-influenced metabolites associated with athletic performance. *Scientific Reports*, *9*(1), 19889. https://doi.org/10.1038/s41598-019-56496-7

Allen, M. S., Greenlees, I., & Jones, M. (2013). Personality in sport: A comprehensive review. *International Review of Sport and Exercise Psychology*, *6*(1), 184–208. https://doi.org/10.1080/1750984X.2013.769614

Althouse, A. D. (2016). Adjust for multiple comparisons? It's not that simple. *The Annals of Thoracic Surgery*, *101*(5), 1644–1645. https://doi.org/10.1016/j.athoracsur.2015.11.024

Alvarez, R., Terrados, N., Ortolano, R., Iglesias-Cubero, G., Reguero, J. R., Batalla, A., Cortina, A., Fernández-García, B., Rodríguez, C., Braga, S., Alvarez, V., & Coto, E. (2000). Genetic variation in the renin-angiotensin system and athletic performance.

European Journal of Applied Physiology, *82*(1–2), 117–120. https://doi.org/10.1007/s004210050660

Andreeva, L., Petrov, L., Mugandani, S. C., Djarova, T., Atanasov, P., & Watson, G. (2014). Angiotensin-converting enzyme genotypes, allele frequency, C-reactive protein, uric acid in female Zulu South African soccer, netball and Bulgarian soccer players. *African Journal for Physical Health Education, Recreation and Dance*, *20*(1), 153–163.

Aro, A. R., Hakonen, A., Hietala, M., Lönnqvist, J., Niemelä, P., Peltonen, L., & Aula, P. (1997). Acceptance of genetic testing in a general population: Age, education and gender differences. *Patient Education and Counseling*, *32*(1–2), 41–49. https://doi.org/10.1016/s0738-3991(97)00061-x

Artells, R., Pruna, R., Dellal, A., & Maffulli, N. (2016). Elastin: A possible genetic biomarker for more severe ligament injuries in elite soccer. a pilot study. *Muscles, Ligaments and Tendons Journal*, *6*(2), 188–192. https://doi.org/10.11138/mltj/2016.6.2.188

Atanasov, P, Djarova, T, Kalinski, M, Petrov, L, Radka Kaneva, Mugandani, S, Watson, G, & Jemni, M. (2015). ACTN3 and AMPD1 polymorphism and genotype combinations in Bulgarian athletes performing wingate test. *Journal of Sports Science*, *3*(1), 1–10. https://doi.org/10.17265/2332-7839/2015.01.001

Attia, J., Ioannidis, J. P. A., Thakkinstian, A., McEvoy, M., Scott, R. J., Minelli, C., Thompson, J., Infante-Rivard, C., & Guyatt, G. (2009). How to use an article about genetic association: B: Are the results of the study valid? *JAMA*, *301*(2), 191–197. https://doi.org/10.1001/jama.2008.946

Ausmees, L., Talts, M., Allik, J., Vainik, U., Sikka, T. T., Nikopensius, T., Esko, T., & Realo, A. (2021). Taking risks to feel excitement: Detailed personality profile and genetic

associations. *European Journal of Personality*, 08902070211019242. https://doi.org/10.1177/08902070211019242

Austin, E. J. (2005). Personality correlates of the broader autism phenotype as assessed by the Autism Spectrum Quotient (AQ). *Personality and Individual Differences*, *38*(2), 451–460. https://doi.org/10.1016/j.paid.2004.04.022

Aziz, A. R., Mukherjee, S., Chia, M. Y. H., & Teh, K. C. (2008). Validity of the running repeated sprint ability test among playing positions and level of competitiveness in trained soccer players. *International Journal of Sports Medicine*, *29*(10), 833–838. https://doi.org/10.1055/s-2008-1038410

Baker, J., Wattie, N., & Schorer, J. (2019). A proposed conceptualization of talent in sport: The first step in a long and winding road. *Psychology of Sport and Exercise*, *43*, 27–33. https://doi.org/10.1016/j.psychsport.2018.12.016

Balestri, M., Calati, R., Serretti, A., & De Ronchi, D. (2014). Genetic modulation of personality traits: A systematic review of the literature. *International Clinical Psychopharmacology*, *29*(1), 1–15. https://doi.org/10.1097/YIC.0b013e328364590b

Baltazar-Martins, G., Gutiérrez-Hellín, J., Aguilar-Navarro, M., Ruiz-Moreno, C., Moreno-Pérez, V., López-Samanes, Á., Domínguez, R., & Del Coso, J. (2020). Effect of ACTN3 genotype on sports performance, exercise-induced muscle damage, and injury epidemiology. *Sports*, *8*(7). https://doi.org/10.3390/sports8070099

Bangsbo, J. (2014). Physiological demands of football. *Sports Science Exchange*, *27*(125), 1–6.

Bangsbo, J., Mohr, M., & Krustrup, P. (2006). Physical and metabolic demands of training and match-play in the elite football player. *Journal of Sports Sciences*, *24*(7), 665–674. https://doi.org/10.1080/02640410500482529

Barbosa, F. B., Cagnin, N. F., Simioni, M., Farias, A. A., Torres, F. R., Molck, M. C., Araujo, T. K., Gil-Da-Silva-Lopes, V. L., Donadi, E. A., & Simões, A. L. (2017). Ancestry informative marker panel to estimate population stratification using genome-wide human array. *Annals of Human Genetics*, *81*(6), 225–233. https://doi.org/10.1111/ahg.12208

Barnes, C., Archer, D. T., Hogg, B., Bush, M., & Bradley, P. S. (2014). The evolution of physical and technical performance parameters in the English Premier League. *International Journal of Sports Medicine*, *35*(13), 1095–1100. https://doi.org/10.1055/s-0034-1375695

Baumert, P., Hall, E. C., & Erskine, R. M. (2019). Chapter Seventeen—The genetic association with exercise-induced muscle damage and muscle injury risk. In D. Barh & I. I. Ahmetov (Eds.), *Sports, Exercise, and Nutritional Genomics* (pp. 375–407). Academic Press. https://doi.org/10.1016/B978-0-12-816193-7.00017-8

Beggs, A. H., Byers, T. J., Knoll, J. H., Boyce, F. M., Bruns, G. A., & Kunkel, L. M. (1992). Cloning and characterization of two human skeletal muscle alpha-actinin genes located on chromosomes 1 and 11. *The Journal of Biological Chemistry*, *267*(13), 9281–9288.

Bell, W., Colley, J. P., Evans, W. D., Darlington, S. E., & Cooper, S.-M. (2012). ACTN3 genotypes of Rugby Union players: Distribution, power output and body composition. *Annals of Human Biology*, *39*(1), 19–27. https://doi.org/10.3109/03014460.2011.632648

Benítez-Sillero, J. de D., Martínez-Aranda, L. M., Sanz-Matesanz, M., & Domínguez-Escribano, M. (2021). Determining factors of psychological performance and differences

among age categories in youth football players. *Sustainability*, *13*(14), Article 14. https://doi.org/10.3390/su13147713

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, *57*(1), 289–300. JSTOR.

Bennett, K. J., Novak, A. R., Pluss, M. A., Coutts, A. J., & Fransen, J. (2020). A multifactorial comparison of Australian youth soccer players' performance characteristics. *International Journal of Sports Science & Coaching*, *15*(1), 17–25. https://doi.org/10.1177/1747954119893174

Berdeli, A., & Cam, F. S. (2009). Prevalence of the angiotensin I converting enzyme gene insertion/deletion polymorphism in a healthy Turkish population. *Biochemical Genetics*, *47*(5–6), 412–420. https://doi.org/10.1007/s10528-009-9240-8

Bergkamp, T. L. G., Niessen, A. S. M., den Hartigh, R. J. R., Frencken, W. G. P., & Meijer, R. R. (2019). Methodological issues in soccer talent identification research. *Sports Medicine*, *49*(9), 1317–1335. https://doi.org/10.1007/s40279-019-01113-w

Biesta, G. (2010). Pragmatism and the philosophical foundations of mixed methods research. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of Mixed Methods in Social and Behavioral Research* (2nd ed., pp. 95–117). Sage Publications.

Bloomfield, J., Polman, R., & O'Donoghue, P. (2007). Physical demands of different positions in FA premier league soccer. *Journal of Sports Science & Medicine*, *6*(1), 63–70.

Bondareva, E. A., Andreev, R. S., Yakushkin, A. V., Parfenteva, O. I., Akimov, E. B., & Sonkin, V. D. (2016). Polymorphism of uncoupling protein genes in football players: Investigation of the functional role. *Fiziologiia Cheloveka*, *42*(6), 70–80.

Bouchard, C. (2011). Overcoming barriers to progress in exercise genomics. *Exercise and Sport Sciences Reviews*, *39*(4), 212–217. https://doi.org/10.1097/JES.0b013e31822643f6

Bouchard, C. (2012). Genomic predictors of trainability. *Experimental Physiology*, *97*(3), 347–352. https://doi.org/10.1113/expphysiol.2011.058735

Bouchard, C. (2015). Exercise genomics—a paradigm shift is needed: A commentary. *British Journal of Sports Medicine*, *49*(23), 1492–1496. https://doi.org/10.1136/bjsports-2015-095294

Bouthenet, M. L., Souil, E., Martres, M. P., Sokoloff, P., Giros, B., & Schwartz, J. C. (1991). Localization of dopamine D3 receptor mRNA in the rat brain using in situ hybridization histochemistry: Comparison with dopamine D2 receptor mRNA. *Brain Research*, *564*(2), 203–219. https://doi.org/10.1016/0006-8993(91)91456-b

Bradley, P. S., Sheldon, W., Wooster, B., Olsen, P., Boanas, P., & Krustrup, P. (2009). High-intensity running in English FA Premier League soccer matches. *Journal of Sports Sciences*, *27*(2), 159–168. https://doi.org/10.1080/02640410802512775

Brazier, J., Antrobus, M., Stebbings, G. K., Day, S. H., Heffernan, S. M., Cross, M. J., & Williams, A. G. (2019). Tendon and ligament injuries in elite rugby: The potential genetic influence. *Sports (Basel, Switzerland)*, *7*(6). https://doi.org/10.3390/sports7060138

Brodde, O.-E. (2008). Beta1- and beta2-adrenoceptor polymorphisms and cardiovascular diseases. *Fundamental & Clinical Pharmacology*, *22*(2), 107–125. https://doi.org/10.1111/j.1472-8206.2007.00557.x

Broos, S., Malisoux, L., Theisen, D., Francaux, M., Deldicque, L., & Thomis, M. A. (2012). Role of alpha-actinin-3 in contractile properties of human single muscle fibers: A case series study in paraplegics. *PLoS ONE*, *7*(11). https://doi.org/10.1371/journal.pone.0049281

Buchheit, M., Mendez-villanueva, A., Simpson, B. M., & Bourdon, P. C. (2010). Repeated-sprint sequences during youth soccer matches. *International Journal of Sports Medicine*, *31*(10), 709–716. https://doi.org/10.1055/s-0030-1261897

Cagliani, R., Fumagalli, M., Pozzoli, U., Riva, S., Comi, G. P., Torri, F., Macciardi, F., Bresolin, N., & Sironi, M. (2009). Diverse evolutionary histories for beta-adrenoreceptor genes in humans. *American Journal of Human Genetics*, *85*(1), 64–75. https://doi.org/10.1016/j.ajhg.2009.06.005

Calvo, M., Rodas, G., Vallejo, M., Estruch, A., Arcas, A., Javierre, C., Viscor, G., & Ventura, J. L. (2002). Heritability of explosive power and anaerobic capacity in humans. *European Journal of Applied Physiology*, *86*(3), 218–225. https://doi.org/10.1007/s004210100522

Camporesi, S., & McNamee, M. J. (2016). Ethics, genetic testing, and athletic talent: Children's best interests, and the right to an open (athletic) future. *Physiological Genomics*, *48*(3), 191–195. https://doi.org/10.1152/physiolgenomics.00104.2015

Carey, A. L., Steinberg, G. R., Macaulay, S. L., Thomas, W. G., Holmes, A. G., Ramm, G., Prelovsek, O., Hohnen-Behrens, C., Watt, M. J., James, D. E., Kemp, B. E., Pedersen, B. K., & Febbraio, M. A. (2006). Interleukin-6 increases insulin-stimulated glucose disposal in humans and glucose uptake and fatty acid oxidation in vitro via AMP-activated protein kinase. *Diabetes*, *55*(10), 2688–2697. https://doi.org/10.2337/db05-1404

Casamichana, D., Castellano, J., & Castagna, C. (2012). Comparing the physical demands of friendly matches and small-sided games in semiprofessional soccer players. *Journal of Strength and Conditioning Research*, *26*(3), 837–843. https://doi.org/10.1519/JSC.0b013e31822a61cf

Cauci, S., Migliozzi, F., Trombetta, C. S., Venuto, I., Saccheri, P., Travan, L., & Chiriacò, G. (2017). Low back pain and FokI (rs2228570) polymorphism of vitamin D receptor in athletes. *BMC Sports Science, Medicine & Rehabilitation*, *9*, 4. https://doi.org/10.1186/s13102-017-0069-x

Chabris, C. F., Lee, J. J., Cesarini, D., Benjamin, D. J., & Laibson, D. I. (2015). The fourth law of behavior genetics. *Current Directions in Psychological Science*, *24*(4), 304–312. https://doi.org/10.1177/0963721415580430

Charlier, R., Caspers, M., Knaeps, S., Mertens, E., Lambrechts, D., Lefevre, J., & Thomis, M. (2017). Limited potential of genetic predisposition scores to predict muscle mass and strength performance in Flemish Caucasians between 19 and 73 years of age. *Physiological Genomics*, *49*(3), 160–166. https://doi.org/10.1152/physiolgenomics.00085.2016

Cheslack-Postava, K., Fallin, M. D., Avramopoulos, D., Connors, S. L., Zimmerman, A. W., Eberhart, C. G., & Newschaffer, C. J. (2007). Beta2-Adrenergic receptor gene variants and risk for autism in the AGRE cohort. *Molecular Psychiatry*, *12*(3), 283–291. https://doi.org/10.1038/sj.mp.4001940

Chiu, L.-L., Wu, Y.-F., Tang, M.-T., Yu, H.-C., Hsieh, L.-L., & Hsieh, S. S.-Y. (2011). ACTN3 genotype and swimming performance in Taiwan. *International Journal of Sports Medicine*, *32*(6), 476–480. https://doi.org/10.1055/s-0030-1263115

Cichon, S., Nöthen, M. M., Erdmann, J., & Propping, P. (1994). Detection of four polymorphic sites in the human dopamine D1 receptor gene (DRD1). *Human Molecular Genetics*, *3*(1), 209. https://doi.org/10.1093/hmg/3.1.209

Cięszczyk, P., Eider, J., Ostanek, M., Arczewska, A., Leońska-Duniec, A., Sawczyn, S., Ficek, K., & Krupecki, K. (2011). Association of the ACTN3 R577X polymorphism in polish power-orientated athletes. *Journal of Human Kinetics*, *28*, 55–61. https://doi.org/10.2478/v10078-011-0022-0

Cieszczyk, P., Krupecki, K., Maciejewska, A., & Sawczuk, M. (2009). The angiotensin converting enzyme gene I/D polymorphism in Polish rowers. *International Journal of Sports Medicine*, *30*(8), 624–627. https://doi.org/10.1055/s-0029-1202825

Cięszczyk, P., Leońska-Duniec, A., Maciejewska-Skrendo, A., Sawczuk, M., Leźnicka, K., Contrò, V., Trybek, G., & Lulińska-Kuklik, E. (2016). Variation in the ace gene in elite polish football players. *Human Movement*, *17*(4), 237–241. https://doi.org/10.1515/humo-2016-0032

Cięszczyk, P., Willard, K., Gronek, P., Zmijewski, P., Trybek, G., Gronek, J., Weber-Rajek, M., Stastny, P., Petr, M., Lulińska-Kuklik, E., Ficek, K., Kemeryte-Riaubiene, E., Maculewicz, E., & September, A. (2017). Are genes encoding proteoglycans really associated with the risk of anterior cruciate ligament rupture? *Biology of Sport*, *34*(2), 97–103. https://doi.org/10.5114/biolsport.2017.64582

Clos, E., Pruna, R., Lundblad, M., Artells, R., & Esquirol Caussa, J. (2019). ACTN3 single nucleotide polymorphism is associated with non-contact musculoskeletal soft-tissue injury incidence in elite professional football players. *Knee Surgery, Sports*

Traumatology, Arthroscopy: Official Journal of the ESSKA, *27*(12), 4055–4061. https://doi.org/10.1007/s00167-019-05381-x

Cocci, P., Pistolesi, L., Guercioni, M., Belli, L., Carli, D., & Palermo, F. A. (2019). Genetic variants and mixed sport disciplines: A comparison among soccer, combat and motorcycle athletes. *Annals of Applied Sport Science*, *7*(1), 1–9. https://doi.org/10.29252/aassjournal.7.1.1

Cochrane, G. D., Sundman, M. H., Hall, E. E., Kostek, M. C., Patel, K., Barnes, K. P., & Ketcham, C. J. (2018). Genetics influence neurocognitive performance at baseline but not concussion history in collegiate student-athletes. *Clinical Journal of Sport Medicine: Official Journal of the Canadian Academy of Sport Medicine*, *28*(2), 125–129. https://doi.org/10.1097/JSM.0000000000000443

Coelho, D. B., Pimenta, E. M., Rosse, I. C., de Castro, B. M., Becker, L. K., de Oliveira, E. C., Carvalho, M. R. S., & Garcia, E. S. (2018). Evidence for a role of ACTN3 R577X polymorphism in football player's career progression. *International Journal of Sports Medicine*, *39*(14), 1088–1093. https://doi.org/10.1055/a-0753-4973

Coelho, D. B., Pimenta, E. M., Rosse, I. C., Veneroso, C., Pussieldi, G. D. A., Becker, L. K., Oliveira, E. C., Carvalho, M. R. S., & Silami-Garcia, E. (2019). Alpha-Actinin-3 R577X polymorphism influences muscle damage and hormonal responses after a soccer game. *Journal of Strength and Conditioning Research*, *33*(10), 2655–2664. https://doi.org/10.1519/JSC.0000000000002575

Coelho, D. B., Pimenta, E., Rosse, I. C., Veneroso, C., Pussieldi, G., Becker, L. K., Carvalho, M.-R., & Silami-Garcia, E. (2016). Angiotensin-converting enzyme (ACE-I/D)

polymorphism frequency in Brazilian soccer players. *Applied Physiology, Nutrition, and Metabolism*, *41*(6), 692–694. https://doi.org/10.1139/apnm-2015-0514

Comings, D. E., Gade, R., Wu, S., Chiu, C., Dietz, G., Muhleman, D., Saucier, G., Ferry, L., Rosenthal, R. J., Lesieur, H. R., Rugle, L. J., & MacMurray, P. (1997). Studies of the potential role of the dopamine D1 receptor gene in addictive behaviors. *Molecular Psychiatry*, *2*(1), 44–56. https://doi.org/10.1038/sj.mp.4000207

Cooper, K. B., Wilson, M. R., & Jones, M. I. (2021). Fast talkers? Investigating the influence of self-talk on mental toughness and finish times in 800-meter runners. *Journal of Applied Sport Psychology*, *33*(5), 491–509. https://doi.org/10.1080/10413200.2020.1735574

Creswell, J. W., & Plano Clark, V. L. (Eds.). (2011). *Designing and conducting mixed methods research* (2nd ed.). Sage Publications.

Cupeiro, R., Pérez-Prieto, R., Amigo, T., Gortázar, P., Redondo, C., & González-Lamuño, D. (2016). Role of the monocarboxylate transporter MCT1 in the uptake of lactate during active recovery. *European Journal of Applied Physiology*, *116*(5), 1005–1010. https://doi.org/10.1007/s00421-016-3365-3

Danser, A. H., Schalekamp, M. A., Bax, W. A., van den Brink, A. M., Saxena, P. R., Riegger, G. A., & Schunkert, H. (1995). Angiotensin-converting enzyme in the human heart. Effect of the deletion/insertion polymorphism. *Circulation*, *92*(6), 1387–1388. https://doi.org/10.1161/01.cir.92.6.1387

Davids, K., & Baker, J. (2007). Genes, environment and sport performance: Why the nature-nurture dualism is no longer relevant. *Sports Medicine*, *37*(11), 961–980. https://doi.org/10.2165/00007256-200737110-00004

de Krom, M., Staal, W. G., Ophoff, R. A., Hendriks, J., Buitelaar, J., Franke, B., de Jonge, M. V., Bolton, P., Collier, D., Curran, S., van Engeland, H., & van Ree, J. M. (2009). A common variant in DRD3 receptor is associated with autism spectrum disorder. *Biological Psychiatry*, *65*(7), 625–630. https://doi.org/10.1016/j.biopsych.2008.09.035

De Moor, M. H. M., Spector, T. D., Cherkas, L. F., Falchi, M., Hottenga, J. J., Boomsma, D. I., & Geus, E. J. C. D. (2007). Genome-wide linkage scan for athlete status in 700 British female DZ twin pairs. *Twin Research and Human Genetics*, *10*(6), 812–820. https://doi.org/10.1375/twin.10.6.812

de Moraes, Í. A. P., Massetti, T., Crocetta, T. B., da Silva, T. D., de Menezes, L. D. C., Monteiro, C. B. de M., & Magalhães, F. H. (2017). Motor learning characterization in people with autism spectrum disorder: A systematic review. *Dementia & Neuropsychologia*, *11*(3), 276–286. https://doi.org/10.1590/1980-57642016dn11-030010

Dellal, A., Chamari, K., Wong, D. P., Ahmaidi, S., Keller, D., Barros, R., Bisciotti, G. N., & Carling, C. (2011). Comparison of physical and technical performance in European soccer match-play: FA Premier League and La Liga. *European Journal of Sport Science*, *11*(1), 51–59. https://doi.org/10.1080/17461391.2010.481334

Delmonico, M. J., Kostek, M. C., Doldo, N. A., Hand, B. D., Walsh, S., Conway, J. M., Carignan, C. R., Roth, S. M., & Hurley, B. F. (2007). Alpha-actinin-3 (ACTN3) R577X polymorphism influences knee extensor peak power response to strength training in older men and women. *The Journals of Gerontology. Series A, Biological Sciences and Medical Sciences*, *62*(2), 206–212. https://doi.org/10.1093/gerona/62.2.206

Deprez, D. N., Fransen, J., Lenoir, M., Philippaerts, R. M., & Vaeyens, R. (2015). A retrospective study on anthropometrical, physical fitness, and motor coordination characteristics that influence dropout, contract status, and first-team playing time in high-level soccer players aged eight to eighteen years. *Journal of Strength and Conditioning Research*, *29*(6), 1692–1704. https://doi.org/10.1519/JSC.0000000000000806

Deveci, S. Ş., Matur, Z., Kesim, Y. Y., Senturk Şentürk, G. G., Sargın-Kurt, G. G., Ugur Uğur, S. A., & Oge Öge, A. E. (2020). Effect of the brain-derived neurotrophic factor gene Val66Met polymorphism on sensory-motor integration during a complex motor learning exercise. *Brain Research*, *1732*, 146652. https://doi.org/10.1016/j.brainres.2020.146652

Di Salvo, V., Baron, R., Tschan, H., Calderon Montero, F. J., Bachl, N., & Pigozzi, F. (2007). Performance characteristics according to playing position in elite soccer. *International Journal of Sports Medicine*, *28*(3), 222–227. https://doi.org/10.1055/s-2006-924294

Di Salvo, V., Gregson, W., Atkinson, G., Tordoff, P., & Drust, B. (2009). Analysis of high intensity activity in Premier League soccer. *International Journal of Sports Medicine*, *30*(3), 205–212. https://doi.org/10.1055/s-0028-1105950

Dinç, N., Yücel, S. B., Taneli, F., & Sayın, M. V. (2016). The effect of the MTHFR C677T mutation on athletic performance and the homocysteine level of soccer players and sedentary individuals. *Journal of Human Kinetics*, *51*, 61–69. https://doi.org/10.1515/hukin-2015-0171

Diogenes, M. E. L., Bezerra, F. F., Cabello, G. M. K., Cabello, P. H., Mendonça, L. M. C., Oliveira Júnior, A. V., & Donangelo, C. M. (2010). Vitamin D receptor gene FokI polymorphisms influence bone mass in adolescent football (soccer) players. *European*

Journal of Applied Physiology, *108*(1), 31–38. https://doi.org/10.1007/s00421-009-1191-6

Dionísio, T. J., Thiengo, C. R., Brozoski, D. T., Dionísio, E. J., Talamoni, G. A., Silva, R. B., Garlet, G. P., Santos, C. F., & Amaral, S. L. (2017). The influence of genetic polymorphisms on performance and cardiac and hemodynamic parameters among Brazilian soccer players. *Applied Physiology, Nutrition, and Metabolism = Physiologie Appliquee, Nutrition Et Metabolisme*, *42*(6), 596–604. https://doi.org/10.1139/apnm-2016-0608

Dodd, K. D., & Newans, T. J. (2018). Talent identification for soccer: Physiological aspects. *Journal of Science and Medicine in Sport*, *21*(10), 1073–1078. https://doi.org/10.1016/j.jsams.2018.01.009

Dohlman, H. G., Thorner, J., Caron, M. G., & Lefkowitz, R. J. (1991). Model systems for the study of seven-transmembrane-segment receptors. *Annual Review of Biochemistry*, *60*, 653–688. https://doi.org/10.1146/annurev.bi.60.070191.003253

Domańska-Senderowska, D., Jastrzębski, Z., Kiszałkiewicz, J., Brzeziański, M., Pastuszak-Lewandoska, D., Radzimiński, Ł., Brzeziańska-Lasota, E., & Jegier, A. (2017). Expression analysis of selected classes of circulating exosomal miRNAs in soccer players as an indicator of adaptation to physical activity. *Biology of Sport*, *34*(4), 331–338. https://doi.org/10.5114/biolsport.2017.69820

Domańska-Senderowska, D., Snochowska, A., Szmigielska, P., Jastrzębski, Z., Jegier, A., Kiszałkiewicz, J., Dróbka, K., Jastrzębska, J., Pastuszak-Lewandoska, D., Cięszczyk, P., Maciejewska-Skrendo, A., Zmijewski, P., & Brzeziańska-Lasota, E. (2018). Analysis of the PPARD gene expression level changes in football players in response to the training

cycle. *Balkan Journal of Medical Genetics: BJMG*, *21*(1), 19–25. https://doi.org/10.2478/bjmg-2018-0008

Druzhevskaya, A. M., Ahmetov, I. I., Astratenkova, I. V., & Rogozkin, V. A. (2008). Association of the ACTN3 R577X polymorphism with power athlete status in Russians. *European Journal of Applied Physiology*, *103*(6), 631–634. https://doi.org/10.1007/s00421-008-0763-1

Dugdale, J. H., Sanders, D., Myers, T., Williams, A. M., & Hunter, A. M. (2020). A case study comparison of objective and subjective evaluation methods of physical qualities in youth soccer players. *Journal of Sports Sciences*, *38*(11–12), 1304–1312. https://doi.org/10.1080/02640414.2020.1766177

Durmic, T. S., Zdravkovic, M. D., Djelic, M. N., Gavrilovic, T. D., Djordjevic Saranovic, S. A., Plavsic, J. N., Mirkovic, S. V., Batinic, D. V., Antic, M. N., Mihailovic, Z. R., Atanasijevic, N. G., Mileusnic, M. J., & Stojkovic, O. V. (2017). Polymorphisms in ACE and ACTN3 genes and blood pressure response to acute exercise in elite male athletes from Serbia. *The Tohoku Journal of Experimental Medicine*, *243*(4), 311–320. https://doi.org/10.1620/tjem.243.311

Dzau, V. J. (1988). Circulating versus local renin-angiotensin system in cardiovascular homeostasis. *Circulation*, *77*(6 Pt 2), I4-13.

Egger, M., Davey Smith, G., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ (Clinical Research Ed.)*, *315*(7109), 629–634. https://doi.org/10.1136/bmj.315.7109.629

Egorova, E. S., Borisova, A. V., Mustafina, L. J., Arkhipova, A. A., Gabbasov, R. T., Druzhevskaya, A. M., Astratenkova, I. V., & Ahmetov, I. I. (2014). The polygenic profile

of Russian football players. *Journal of Sports Sciences*, *32*(13), 1286–1293. https://doi.org/10.1080/02640414.2014.898853

Ehlert, T., Simon, P., & Moser, D. A. (2013). Epigenetics in sports. *Sports Medicine*, *43*(2), 93–110. https://doi.org/10.1007/s40279-012-0012-y

Eider, J., Cieszczyk, P., Leońska-Duniec, A., Maciejewska, A., Sawczuk, M., Ficek, K., & Kotarska, K. (2013). Association of the 174 G/C polymorphism of the IL6 gene in Polish power-orientated athletes. *The Journal of Sports Medicine and Physical Fitness*, *53*(1), 88–92.

Eirale, C., Tol, J. L., Farooq, A., Smiley, F., & Chalabi, H. (2013). Low injury rate strongly correlates with team success in Qatari professional football. *British Journal of Sports Medicine*, *47*(12), 807–808. https://doi.org/10.1136/bjsports-2012-091040

Ekstrand, J., Hägglund, M., & Waldén, M. (2011). Injury incidence and injury patterns in professional football: The UEFA injury study. *British Journal of Sports Medicine*, *45*(7), 553–558. https://doi.org/10.1136/bjsm.2009.060582

Eliakim, E., Morgulev, E., Lidor, R., & Meckel, Y. (2020). Estimation of injury costs: Financial damage of English Premier League teams' underachievement due to injuries. *BMJ Open Sport & Exercise Medicine*, *6*(1), e000675. https://doi.org/10.1136/bmjsem-2019-000675

Emmonds, S., Till, K., Jones, B., Mellis, M., & Pears, M. (2016). Anthropometric, speed and endurance characteristics of English academy soccer players: Do they influence obtaining a professional contract at 18 years of age? *International Journal of Sports Science & Coaching*, *11*(2), 212–218. https://doi.org/10.1177/1747954116637154

Erdös, E. G., & Skidgel, R. A. (1987). The angiotensin I-converting enzyme. *Laboratory Investigation; a Journal of Technical Methods and Pathology*, *56*(4), 345–348.

Erskine, R. M., Williams, A. G., Jones, D. A., Stewart, C. E., & Degens, H. (2014). The individual and combined influence of ACE and ACTN3 genotypes on muscle phenotypes before and after strength training. *Scandinavian Journal of Medicine & Science in Sports*, *24*(4), 642–648. https://doi.org/10.1111/sms.12055

Eynon, N., Banting, L. K., Ruiz, J. R., Cieszczyk, P., Dyatlov, D. A., Maciejewska-Karlowska, A., Sawczuk, M., Pushkarev, V. P., Kulikov, L. M., Pushkarev, E. D., Femia, P., Stepto, N. K., Bishop, D. J., & Lucia, A. (2014). ACTN3 R577X polymorphism and team-sport performance: A study involving three European cohorts. *Journal of Science and Medicine in Sport*, *17*(1), 102–106. https://doi.org/10.1016/j.jsams.2013.02.005

Eynon, N., Duarte, J. A., Oliveira, J., Sagiv, M., Yamin, C., Meckel, Y., Sagiv, M., & Goldhammer, E. (2009). ACTN3 R577X polymorphism and Israeli top-level athletes. *International Journal of Sports Medicine*, *30*(9), 695–698. https://doi.org/10.1055/s-0029-1220731

Eynon, N., Ruiz, J. R., Oliveira, J., Duarte, J. A., Birk, R., & Lucia, A. (2011). Genes and elite athletes: A roadmap for future research. *The Journal of Physiology*, *589*(13), 3063–3070. https://doi.org/10.1113/jphysiol.2011.207035

Eynon, N., Ruiz, J. R., Yvert, T., Santiago, C., Gómez-Gallego, F., Lucia, A., & Birk, R. (2012). The C allele in NOS3 -786 T/C polymorphism is associated with elite soccer player's status. *International Journal of Sports Medicine*, *33*(7), 521–524. https://doi.org/10.1055/s-0032-1306337

Falahati, A., & Arazi, H. (2019). Association of ACE gene polymorphism with cardiovascular determinants of trained and untrained Iranian men. *Genes and Environment: The Official Journal of the Japanese Environmental Mutagen Society*, *41*, 8. https://doi.org/10.1186/s41021-019-0126-7

Fatini, C., Guazzelli, R., Manetti, P., Battaglini, B., Gensini, F., Vono, R., Toncelli, L., Zilli, P., Capalbo, A., Abbate, R., Gensini, G. F., & Galanti, G. (2000). RAS genes influence exercise-induced left ventricular hypertrophy: An elite athletes study. *Medicine and Science in Sports and Exercise*, *32*(11), 1868–1872. https://doi.org/10.1097/00005768-200011000-00008

Fazio, V. C., Lovell, M. R., Pardini, J. E., & Collins, M. W. (2007). The relation between post concussion symptoms and neurocognitive performance in concussed athletes. *NeuroRehabilitation*, *22*(3), 207–216.

Febbraio, M. A., & Pedersen, B. K. (2002). Muscle-derived interleukin-6: Mechanisms for activation and possible biological roles. *FASEB Journal: Official Publication of the Federation of American Societies for Experimental Biology*, *16*(11), 1335–1347. https://doi.org/10.1096/fj.01-0876rev

Ficek, K., Cieszczyk, P., Kaczmarczyk, M., Maciejewska-Karłowska, A., Sawczuk, M., Cholewinski, J., Leonska-Duniec, A., Stepien-Slodkowska, M., Zarebska, A., Stepto, N. K., Bishop, D. J., & Eynon, N. (2013). Gene variants within the COL1A1 gene are associated with reduced anterior cruciate ligament injury in professional soccer players. *Journal of Science and Medicine in Sport*, *16*(5), 396–400. https://doi.org/10.1016/j.jsams.2012.10.004

Ficek, K., Stepien-Slodkowska, M., Kaczmarczyk, M., Maciejewska-Karlowska, A., Sawczuk, M., Cholewinski, J., Leonska-Duniec, A., Zarebska, A., Cieszczyk, P., & Zmijewski, P. (2014). Does the A9285G polymorphism in Collagen Type XII α1 gene associate with the risk of anterior cruciate ligament ruptures? *Balkan Journal of Medical Genetics: BJMG*, *17*(1), 41–46. https://doi.org/10.2478/bjmg-2014-0022

Filonzi, L., Franchini, N., Vaghi, M., Chiesa, S., & Marzano, F. N. (2015). The potential role of myostatin and neurotransmission genes in elite sport performances. *Journal of Biosciences*, *40*(3), 531–537. https://doi.org/10.1007/s12038-015-9542-4

Fisher, R. A. (1918). The correlation between relatives on the supposition of mendelian inheritance. *Earth and Environmental Science Transactions of The Royal Society of Edinburgh*, *52*(2), 399–433. https://doi.org/10.1017/S0080456800012163

Fishman, D., Faulds, G., Jeffery, R., Mohamed-Ali, V., Yudkin, J. S., Humphries, S., & Woo, P. (1998). 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. *Journal of Clinical Investigation*, *102*(7), 1369–1376.

Flueck, M., Vaughan, D., & Westerblad, H. (2010). Linking genes with exercise: Where is the cut-off? *European Journal of Applied Physiology*, *110*(6), 1095–1098. https://doi.org/10.1007/s00421-010-1662-9

Flynn, R. K., Pedersen, C. L., Birmingham, T. B., Kirkley, A., Jackowski, D., & Fowler, P. J. (2005). The familial predisposition toward tearing the anterior cruciate ligament: A case control study. *The American Journal of Sports Medicine*, *33*(1), 23–28. https://doi.org/10.1177/0363546504265678

Forsdyke, D., Smith, A., Jones, M., & Gledhill, A. (2016). Psychosocial factors associated with outcomes of sports injury rehabilitation in competitive athletes: A mixed studies systematic review. *British Journal of Sports Medicine*, *50*(9), 537–544. https://doi.org/10.1136/bjsports-2015-094850

Forsman, H., Blomqvist, M., Davids, K., Liukkonen, J., & Konttinen, N. (2016). Identifying technical, physiological, tactical and psychological characteristics that contribute to career progression in soccer. *International Journal of Sports Science & Coaching*, *11*(4), 505–513. https://doi.org/10.1177/1747954116655051

Fox, P. W., Hershberger, S. L., & Bouchard, T. J. (1996). Genetic and environmental contributions to the acquisition of a motor skill. *Nature*, *384*(6607), Article 6607. https://doi.org/10.1038/384356a0

Frank, C. B. (2004). Ligament structure, physiology and function. *Journal of Musculoskeletal & Neuronal Interactions*, *4*(2), 199–201.

Freedman, M. L., Reich, D., Penney, K. L., McDonald, G. J., Mignault, A. A., Patterson, N., Gabriel, S. B., Topol, E. J., Smoller, J. W., Pato, C. N., Pato, M. T., Petryshen, T. L., Kolonel, L. N., Lander, E. S., Sklar, P., Henderson, B., Hirschhorn, J. N., & Altshuler, D. (2004). Assessing the impact of population stratification on genetic association studies. *Nature Genetics*, *36*(4), 388–393. https://doi.org/10.1038/ng1333

Fritsch, B., Reis, J., Martinowich, K., Schambra, H. M., Ji, Y., Cohen, L. G., & Lu, B. (2010). Direct current stimulation promotes BDNF-dependent synaptic plasticity: Potential implications for motor learning. *Neuron*, *66*(2), 198–204. https://doi.org/10.1016/j.neuron.2010.03.035

Fuhre, J., Øygard, A., & Sæther, S. A. (2022). Coaches' criteria for talent identification of youth male soccer players. *Sports*, *10*(2), 14. https://doi.org/10.3390/sports10020014

Fuller, C. W., Ekstrand, J., Junge, A., Andersen, T. E., Bahr, R., Dvorak, J., Hägglund, M., McCrory, P., & Meeuwisse, W. H. (2006). Consensus statement on injury definitions and data collection procedures in studies of football (soccer) injuries. *British Journal of Sports Medicine*, *40*(3), 193–201. https://doi.org/10.1136/bjsm.2005.025270

Galeandro, V., Notarnicola, A., Bianco, A., Tafuri, S., Russo, L., Pesce, V., Moretti, B., & Petruzzella, V. (2017). ACTN3/ACE genotypes and mitochondrial genome in professional soccer players performance. *Journal of Biological Regulators and Homeostatic Agents*, *31*(1), 207–213.

Garatachea, N., Verde, Z., Santos-Lozano, A., Yvert, T., Rodriguez-Romo, G., Sarasa, F. J., Hernández-Sánchez, S., Santiago, C., & Lucia, A. (2014). ACTN3 R577X polymorphism and explosive leg-muscle power in elite basketball players. *International Journal of Sports Physiology and Performance*, *9*(2), 226–232. https://doi.org/10.1123/ijspp.2012-0331

Garton, F. C., Seto, J. T., Quinlan, K. G. R., Yang, N., Houweling, P. J., & North, K. N. (2014). α-Actinin-3 deficiency alters muscle adaptation in response to denervation and immobilization. *Human Molecular Genetics*, *23*(7), 1879–1893. https://doi.org/10.1093/hmg/ddt580

Gayagay, G., Yu, B., Hambly, B., Boston, T., Hahn, A., Celermajer, D. S., & Trent, R. J. (1998). Elite endurance athletes and the ACE I allele—The role of genes in athletic performance. *Human Genetics*, *103*(1), 48–50. https://doi.org/10.1007/s004390050781

Gentil, P., Pereira, R. W., Leite, T. K. M., & Bottaro, M. (2011). ACTN3 R577X polymorphism and neuromuscular response to resistance training. *Journal of Sports Science & Medicine*, *10*(2), 393–399.

Georgiades, E., Klissouras, V., Baulch, J., Wang, G., & Pitsiladis, Y. (2017). Why nature prevails over nurture in the making of the elite athlete. *BMC Genomics*, *18*(S8), 835. https://doi.org/10.1186/s12864-017-4190-8

Gil, S. M., Zabala-Lili, J., Bidaurrazaga-Letona, I., Aduna, B., Lekue, J. A., Santos-Concejero, J., & Granados, C. (2014). Talent identification and selection process of outfield players and goalkeepers in a professional soccer club. *Journal of Sports Sciences*, *32*(20), 1931–1939. https://doi.org/10.1080/02640414.2014.964290

Gil, S., Ruiz, F., Irazusta, A., Gil, J., & Irazusta, J. (2007). Selection of young soccer players in terms of anthropometric and physiological factors. *The Journal of Sports Medicine and Physical Fitness*, *47*(1), 25–32.

Gill, J., Merchant-Borna, K., Lee, H., Livingston, W. S., Olivera, A., Cashion, A., Wang, D., & Bazarian, J. J. (2016). Sports-related concussion results in differential expression of nuclear factor-κB pathway genes in peripheral blood during the acute and subacute periods. *The Journal of Head Trauma Rehabilitation*, *31*(4), 269–276. https://doi.org/10.1097/HTR.0000000000000191

Gineviciene, V., Jakaitiene, A., Tubelis, L., & Kucinskas, V. (2014). Variation in the ACE, PPARGC1A and PPARA genes in Lithuanian football players. *European Journal of Sport Science*, *14*(1), S289-295. https://doi.org/10.1080/17461391.2012.691117

Ginevičienė, V., Pranckevičienė, E., Milašius, K., & Kučinskas, V. (2010). Relating fitness phenotypes to genotypes in Lithuanian elite athletes. *Acta Medica Lituanica*, *17*(1), 1–10. https://doi.org/10.2478/v10140-010-0001-0

Gledhill, A., Harwood, C., & Forsdyke, D. (2017). Psychosocial factors associated with talent development in football: A systematic review. *Psychology of Sport and Exercise*, *31*, 93–112. https://doi.org/10.1016/j.psychsport.2017.04.002

Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & Gough, H. G. (2006). The international personality item pool and the future of public-domain personality measures. *Journal of Research in Personality*, *40*(1), 84–96. https://doi.org/10.1016/j.jrp.2005.08.007

Goldkuhl, G. (2012). Pragmatism vs interpretivism in qualitative information systems research. *European Journal of Information Systems*, *21*(2), 135–146. https://doi.org/10.1057/ejis.2011.54

Goles, T., & Hirschheim, R. (2000). The paradigm is dead, the paradigm is dead…long live the paradigm: The legacy of Burrell and Morgan. *Omega*, *28*(3), 249–268. https://doi.org/10.1016/S0305-0483(99)00042-0

Gómez, M.-Á., Mitrotasios, M., Armatas, V., & Lago-Peñas, C. (2018). Analysis of playing styles according to team quality and match location in Greek professional soccer. *International Journal of Performance Analysis in Sport*, *18*(6), 986–997. https://doi.org/10.1080/24748668.2018.1539382

Gonaus, C., & Müller, E. (2012). Using physiological data to predict future career progression in 14- to 17-year-old Austrian soccer academy players. *Journal of Sports Sciences*, *30*(15), 1673–1682. https://doi.org/10.1080/02640414.2012.713980

Gonzalez-Lopez, E., & Vrana, K. E. (2020). Dopamine beta-hydroxylase and its genetic variants in human health and disease. *Journal of Neurochemistry*, *152*(2), 157–181. https://doi.org/10.1111/jnc.14893

Goodlin, G. T., Roos, A. K., Roos, T. R., Hawkins, C., Beache, S., Baur, S., & Kim, S. K. (2014). Applying personal genetic data to injury risk assessment in athletes. *PloS One*, *10*(4), e0122676. https://doi.org/10.1371/journal.pone.0122676

Goodlin, G. T., Roos, T. R., Roos, A. K., & Kim, S. K. (2015). The dawning age of genetic testing for sports injuries. *Clinical Journal of Sport Medicine: Official Journal of the Canadian Academy of Sport Medicine*, *25*(1), 1–5. https://doi.org/10.1097/JSM.0000000000000158

Gu, Q. (2002). Neuromodulatory transmitter systems in the cortex and their role in cortical plasticity. *Neuroscience*, *111*(4), 815–835. https://doi.org/10.1016/s0306-4522(02)00026-x

Gucciardi, D. F., Hanton, S., Gordon, S., Mallett, C. J., & Temby, P. (2015). The concept of mental toughness: Tests of dimensionality, nomological network, and traitness. *Journal of Personality*, *83*(1), 26–44. https://doi.org/10.1111/jopy.12079

Guilherme, J. P. L. F., Bertuzzi, R., Lima-Silva, A. E., Pereira, A. da C., & Lancha Junior, A. H. (2018). Analysis of sports-relevant polymorphisms in a large Brazilian cohort of top-level athletes. *Annals of Human Genetics*, *82*(5), 254–264. https://doi.org/10.1111/ahg.12248

Guilherme, J. P. L. F., & Lancha, A. H. (2020). Total genotype score and athletic status: An exploratory cross-sectional study of a Brazilian athlete cohort. *Annals of Human Genetics*, *84*(2), 141–150. https://doi.org/10.1111/ahg.12353

Guilherme, J. P. L. F., Tritto, A. C. C., North, K. N., Lancha Junior, A. H., & Artioli, G. G. (2014). Genetics and sport performance: Current challenges and directions to the future. *Revista Brasileira de Educação Física e Esporte*, *28*(1), 177–193. https://doi.org/10.1590/S1807-55092014000100177

Gurevich, E. V., & Joyce, J. N. (1999). Distribution of dopamine D3 receptor expressing neurons in the human forebrain: Comparison with D2 receptor expressing neurons. *Neuropsychopharmacology: Official Publication of the American College of Neuropsychopharmacology*, *20*(1), 60–80. https://doi.org/10.1016/S0893-133X(98)00066-9

Haff, G. G. (2019). *Will the genetic screening of athletes change sport as we know it?* The Conversation. http://theconversation.com/will-the-genetic-screening-of-athletes-change-sport-as-we-know-it-122781

Hägglund, M., Waldén, M., Magnusson, H., Kristenson, K., Bengtsson, H., & Ekstrand, J. (2013). Injuries affect team performance negatively in professional football: An 11-year follow-up of the UEFA Champions League injury study. *British Journal of Sports Medicine*, *47*(12), 738–742. https://doi.org/10.1136/bjsports-2013-092215

Hall, E. C. R., Baumert, P., Larruskain, J., Gil, S. M., Lekue, J. A., Rienzi, E., Moreno, S., Tannure, M., Murtagh, C. F., Ade, J. D., Squires, P., Orme, P., Anderson, L., Brownlee, T. E., Whitworth-Turner, C. M., Morton, J. P., Drust, B., Williams, A. G., & Erskine, R. M. (2022). The genetic association with injury risk in male academy soccer players depends on maturity status. *Scandinavian Journal of Medicine & Science in Sports*, *32*(2), 338–350. https://doi.org/10.1111/sms.14077

Hambrick, D. Z., Burgoyne, A. P., Macnamara, B. N., & Ullén, F. (2018). Toward a multifactorial model of expertise: Beyond born versus made. *Annals of the New York Academy of Sciences*. https://doi.org/10.1111/nyas.13586

Harvie, P., Ostlere, S. J., Teh, J., McNally, E. G., Clipsham, K., Burston, B. J., Pollard, T. C. B., & Carr, A. J. (2004). Genetic influences in the aetiology of tears of the rotator cuff. Sibling risk of a full-thickness tear. *The Journal of Bone and Joint Surgery. British Volume*, *86*(5), 696–700. https://doi.org/10.1302/0301-620x.86b5.14747

Hauber, W. (1998). Involvement of basal ganglia transmitter systems in movement initiation. *Progress in Neurobiology*, *56*(5), 507–540. https://doi.org/10.1016/s0301-0082(98)00041-0

Haugen, T., Tønnessen, E., Hisdal, J., & Seiler, S. (2014). The role and development of sprinting speed in soccer. *International Journal of Sports Physiology and Performance*, *9*(3), 432–441. https://doi.org/10.1123/ijspp.2013-0121

Head, S. I., Chan, S., Houweling, P. J., Quinlan, K. G. R., Murphy, R., Wagner, S., Friedrich, O., & North, K. N. (2015). Altered Ca2+ kinetics associated with α-actinin-3 deficiency may explain positive selection for ACTN3 null allele in human evolution. *PLoS Genetics*, *11*(2), e1004862. https://doi.org/10.1371/journal.pgen.1004862

Heffernan, S. M., Kilduff, L. P., Day, S. H., Pitsiladis, Y. P., & Williams, A. G. (2015). Genomics in rugby union: A review and future prospects. *European Journal of Sport Science*, *15*(6), 460–468. https://doi.org/10.1080/17461391.2015.1023222

Heffernan, S. M., Kilduff, L. P., Erskine, R. M., Day, S. H., McPhee, J. S., McMahon, G. E., Stebbings, G. K., Neale, J. P. H., Lockey, S. J., Ribbans, W. J., Cook, C. J., Vance, B., Raleigh, S. M., Roberts, C., Bennett, M. A., Wang, G., Collins, M., Pitsiladis, Y. P., &

Williams, A. G. (2016). Association of ACTN3 R577X but not ACE I/D gene variants with elite rugby union player status and playing position. *Physiological Genomics*, *48*(3), 196–201. https://doi.org/10.1152/physiolgenomics.00107.2015

Heinegård, D. (2009). Proteoglycans and more – from molecules to biology. *International Journal of Experimental Pathology*, *90*(6), 575–586. https://doi.org/10.1111/j.1365-2613.2009.00695.x

Hendry, D. T., Williams, A. M., & Hodges, N. J. (2018). Coach ratings of skills and their relations to practice, play and successful transitions from youth-elite to adult-professional status in soccer. *Journal of Sports Sciences*, *36*(17), 2009–2017. https://doi.org/10.1080/02640414.2018.1432236

Herbert, A. J., Williams, A. G., Hennis, P. J., Erskine, R. M., Sale, C., Day, S. H., & Stebbings, G. K. (2019). The interactions of physical activity, exercise and genetics and their associations with bone mineral density: Implications for injury risk in elite athletes. *European Journal of Applied Physiology*, *119*(1), 29–47. https://doi.org/10.1007/s00421-018-4007-8

Holdys, J., Kryściak, J., Stanisławski, D., & Gronek, P. (2011). ACE I/D gene polymorphism in athletes of various sports disciplines. *Human Movement*, *12*(3), 223–231. https://doi.org/10.2478/v10038-011-0022-x

Holmes, E. (2018). *In the blood: How DNAFit is using gene testing to prepare Egypt for their World Cup return*. https://www.sportspromedia.com/interviews/dnafit-gene-testing-egypt-world-cup

Honarpour, A., Mohseni, M., Ghavidel Hajiagha, S., Irani, S., & Najmabadi, H. (2017). Investigation of the relationship between a genetic polymorphism in ACTN3 and elite

sport performance among Iranian soccer players. *Iranian Rehabilitation Journal*, *15*(2), 149–154. https://doi.org/10.18869/nrip.irj.15.2.149

Höner, O., Leyhr, D., & Kelava, A. (2017). The influence of speed abilities and technical skills in early adolescence on adult success in soccer: A long-term prospective analysis using ANOVA and SEM approaches. *PloS One*, *12*(8), e0182211. https://doi.org/10.1371/journal.pone.0182211

Höner, O., Murr, D., Larkin, P., Schreiner, R., & Leyhr, D. (2021). Nationwide subjective and objective assessments of potential talent predictors in elite youth soccer: An investigation of prognostic validity in a prospective study. *Frontiers in Sports and Active Living*, *3*, 638227. https://doi.org/10.3389/fspor.2021.638227

Hong, E. P., & Park, J. W. (2012). Sample Size and statistical power calculation in genetic association studies. *Genomics & Informatics*, *10*(2), 117–122. https://doi.org/10.5808/GI.2012.10.2.117

Hopkins, W. G., Marshall, S. W., Batterham, A. M., & Hanin, J. (2009). Progressive statistics for studies in sports medicine and exercise science. *Medicine and Science in Sports and Exercise*, *41*(1), 3–13. https://doi.org/10.1249/MSS.0b013e31818cb278

Hoppeler, H., Baum, O., Lurman, G., & Mueller, M. (2011). Molecular mechanisms of muscle plasticity with exercise. *Comprehensive Physiology*, *1*(3), 1383–1412. https://doi.org/10.1002/cphy.c100042

Horsburgh, V. A., Schermer, J. A., Veselka, L., & Vernon, P. A. (2009). A behavioural genetic study of mental toughness and personality. *Personality and Individual Differences*, *46*(2), 100–105. https://doi.org/10.1016/j.paid.2008.09.009

Hosmer, D., Stanley, L., & Rodney, S. (2013). *Applied Logistic Regression*. John Wiley & Sons.

Hughes, D. C., Day, S. H., Ahmetov, I. I., & Williams, A. G. (2011). Genetics of muscle strength and power: Polygenic profile similarity limits skeletal muscle performance. *Journal of Sports Sciences*, *29*(13), 1425–1434. https://doi.org/10.1080/02640414.2011.597773

Huijgen, B. C. H., Elferink-Gemser, M. T., Post, W. J., & Visscher, C. (2009). Soccer skill development in professionals. *International Journal of Sports Medicine*, *30*(8), 585–591. https://doi.org/10.1055/s-0029-1202354

Ioannidis, J. P., Ntzani, E. E., Trikalinos, T. A., & Contopoulos-Ioannidis, D. G. (2001). Replication validity of genetic association studies. *Nature Genetics*, *29*(3), 306–309. https://doi.org/10.1038/ng749

Ivarsson, A., Kilhage-Persson, A., Martindale, R., Priestley, D., Huijgen, B., Ardern, C., & McCall, A. (2020). Psychological factors and future performance of football players: A systematic review with meta-analysis. *Journal of Science and Medicine in Sport*, *23*(4), 415–420. https://doi.org/10.1016/j.jsams.2019.10.021

Iwao-Koizumi, K., Ota, T., Hayashida, M., Yonetani, Y., Nakata, K., Kinoshita, K., Koyanagi, Y., & Murata, S. (2015). *The ACTN3 Gene is a potential biomarker for the risk of non-contact sports injury in female athletes*. https://doi.org/10.4172/2155-9929.S6-002

Jacob, Y., Chivers, P., & Anderton, R. S. (2019). Genetic predictors of match performance in sub-elite Australian football players: A pilot study. *Journal of Exercise Science and Fitness*, *17*(2), 41–46. https://doi.org/10.1016/j.jesf.2018.10.007

Jacob, Y., Cripps, A., Evans, T., Chivers, P. T., Joyce, C., & Anderton, R. S. (2018). Identification of genetic markers for skill and athleticism in sub-elite Australian football players: A pilot study. *The Journal of Sports Medicine and Physical Fitness*, *58*(3), 241–248. https://doi.org/10.23736/S0022-4707.16.06647-0

Jeong, T.-S., Bartlett, J. D., Joo, C.-H., Louhelainen, J., Close, G. L., Morton, J. P., & Drust, B. (2015). Acute simulated soccer-specific training increases PGC-1α mRNA expression in human skeletal muscle. *Journal of Sports Sciences*, *33*(14), 1493–1503. https://doi.org/10.1080/02640414.2014.992937

Jeremic, D., Macuzic, I. Z., Vulovic, M., Stevanovic, J., Radovanovic, D., Varjacic, V., & Djordjevic, D. (2019). ACE/ACTN3 genetic polymorphisms and athletic performance of female soccer players`. *Revista Brasileira de Medicina Do Esporte*, *25*(1), 35–39. https://doi.org/10.1590/1517-869220192501187684

John, R., Dhillon, M. S., Sharma, S., Prabhakar, S., & Bhandari, M. (2016). Is there a genetic predisposition to anterior cruciate ligament tear? A systematic review. *The American Journal of Sports Medicine*, *44*(12), 3262–3269. https://doi.org/10.1177/0363546515624467

Johnson, M. (2006). Molecular mechanisms of beta(2)-adrenergic receptor function, response, and regulation. *The Journal of Allergy and Clinical Immunology*, *117*(1), 18–24; quiz 25. https://doi.org/10.1016/j.jaci.2005.11.012

Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher*, *33*(7), 14–26. https://doi.org/10.3102/0013189X033007014

Jones, M. I., & Parker, J. K. (2018). Mindfulness mediates the relationship between mental toughness and pain catastrophizing in cyclists. *European Journal of Sport Science*, *18*(6), 872–881. https://doi.org/10.1080/17461391.2018.1478450

Jones, M. I., & Parker, J. K. (2019). An analysis of the size and direction of the association between mental toughness and Olympic distance personal best triathlon times. *Journal of Sport and Health Science*, *8*(1), 71–76. https://doi.org/10.1016/j.jshs.2017.05.005

Jones, N., Kiely, J., Suraci, B., Collins, D. J., de Lorenzo, D., Pickering, C., & Grimaldi, K. A. (2016). A genetic-based algorithm for personalized resistance training. *Biology of Sport*, *33*(2), 117–126. https://doi.org/10.5604/20831862.1198210

Juffer, P., Furrer, R., González-Freire, M., Santiago, C., Verde, Z., Serratosa, L., Morate, F. J., Rubio, J. C., Martin, M. A., Ruiz, J. R., Arenas, J., Gómez-Gallego, F., & Lucia, A. (2009). Genotype distributions in top-level soccer players: A role for ACE? *International Journal of Sports Medicine*, *30*(5), 387–392. https://doi.org/10.1055/s-0028-1105931

Kalapotharakos, V. I., Strimpakos, N., Vithoulka, I., Karvounidis, C., Diamantopoulos, K., & Kapreli, E. (2006). Physiological characteristics of elite professional soccer teams of different ranking. *The Journal of Sports Medicine and Physical Fitness*, *46*(4), 515–519.

Karanikolou, A., Wang, G., & Pitsiladis, Y. (2017). Letter to the editor: A genetic-based algorithm for personalized resistance training. *Biology of Sport*, *34*(1), 31–33. https://doi.org/10.5114/biolsport.2017.63385

Karlsson Linnér, R., Biroli, P., Kong, E., Meddens, S. F. W., Wedow, R., Fontana, M. A., Lebreton, M., Tino, S. P., Abdellaoui, A., Hammerschlag, A. R., Nivard, M. G., Okbay, A., Rietveld, C. A., Timshel, P. N., Trzaskowski, M., Vlaming, R. de, Zünd, C. L., Bao, Y., Buzdugan, L., … Beauchamp, J. P. (2019). Genome-wide association analyses of risk

tolerance and risky behaviors in over 1 million individuals identify hundreds of loci and shared genetic influences. *Nature Genetics*, *51*(2), 245–257. https://doi.org/10.1038/s41588-018-0309-3

Keller, B. S., Raynor, A. J., Bruce, L., & Iredale, F. (2016). Technical attributes of Australian youth soccer players: Implications for talent identification. *International Journal of Sports Science & Coaching*, *11*(6), 819–824. https://doi.org/10.1177/1747954116676108

Kelly, A. L., & Williams, C. A. (2020). Physical characteristics and the talent identification and development processes in male youth soccer: A narrative review. *Strength & Conditioning Journal*, *42*(6), 15–34. https://doi.org/10.1519/SSC.0000000000000576

Kelly, A. L., Wilson, M. R., Jackson, D. T., Goldman, D. E., Turnnidge, J., Côté, J., & Williams, C. A. (2021). A multidisciplinary investigation into 'playing-up' in academy football according to age phase. *Journal of Sports Sciences*, *39*(8), 854–864. https://doi.org/10.1080/02640414.2020.1848117

Kelly, A. L., Wilson, M. R., Jackson, D. T., Turnnidge, J., & Williams, C. A. (2021). Speed of thought and speed of feet: Examining perceptual-cognitive expertise and physical performance in an English football academy. *Journal of Science in Sport and Exercise*, *3*(1), 88–97. https://doi.org/10.1007/s42978-020-00081-2

Kelly, A. L., Wilson, M. R., Jackson, D. T., & Williams, C. A. (2020). Technical testing and match analysis statistics as part of the talent development process in an English football academy. *International Journal of Performance Analysis in Sport*, *20*(6), 1035–1051. https://doi.org/10.1080/24748668.2020.1824865

Kikuchi, N., & Nakazato, K. (2015). Effective utilization of genetic information for athletes and coaches: Focus on ACTN3 R577X polymorphism. *Journal of Exercise Nutrition & Biochemistry*, *19*(3), 157–164. https://doi.org/10.5717/jenb.2015.15093001

Kim, E., Goren, A., & Ast, G. (2008). Alternative splicing: Current perspectives. *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology*, *30*(1), 38–47. https://doi.org/10.1002/bies.20692

Kim, S. M., Oh, S. D., Jung, I. G., Lee, J., Sim, Y. J., Lee, J. K., & Kang, B. Y. (2010). Distribution of the Trp64Arg polymorphism in the ß 3-adrenergic receptor gene in athletes and its influence on cardiovascular function. *Kardiologia Polska*, *68*(8), 920–926.

Kirkland, A., & O'Sullivan, M. (2018). There is no such thing as an international elite under-9 soccer player. *Journal of Sports Science & Medicine*, *17*(4), 686–688.

Kitazawa, H., Hasegawa, K., Aruga, D., & Tanaka, M. (2021). Potential genetic contributions of the central nervous system to a predisposition to elite athletic traits: State-of-the-art and future perspectives. *Genes*, *12*(3). https://doi.org/10.3390/genes12030371

Kleim, J. A., Chan, S., Pringle, E., Schallert, K., Procaccio, V., Jimenez, R., & Cramer, S. C. (2006). BDNF val66met polymorphism is associated with modified experience-dependent plasticity in human motor cortex. *Nature Neuroscience*, *9*(6), 735–737. https://doi.org/10.1038/nn1699

Koku, F. E., Karamızrak, S. O., Çiftçi, A. S., Taşlıdere, H., Durmaz, B., & Çoğulu, Ö. (2019). The relationship between ACTN3 R577X gene polymorphism and physical performance in amateur soccer players and sedentary individuals. *Biology of Sport*, *36*(1), 9–16. https://doi.org/10.5114/biolsport.2018.78900

Koopmann, T., Faber, I., Baker, J., & Schorer, J. (2020). Assessing technical skills in talented youth athletes: A systematic review. *Sports Medicine (Auckland, N.Z.)*, *50*(9), 1593–1611. https://doi.org/10.1007/s40279-020-01299-4

Kozlovskaia, M., Vlahovich, N., Ashton, K. J., & Hughes, D. C. (2017). Biomedical risk factors of achilles tendinopathy in physically active people: A systematic review. *Sports Medicine - Open*, *3*. https://doi.org/10.1186/s40798-017-0087-y

Kraemer, R., Wuerfel, W., Lorenzen, J., Busche, M., Vogt, P. M., & Knobloch, K. (2012). Analysis of hereditary and medical risk factors in Achilles tendinopathy and Achilles tendon ruptures: A matched pair analysis. *Archives of Orthopaedic and Trauma Surgery*, *132*(6), 847–853. https://doi.org/10.1007/s00402-012-1476-9

Kraemer, W. J., & Ratamess, N. A. (2004). Fundamentals of resistance training: Progression and exercise prescription. *Medicine and Science in Sports and Exercise*, *36*(4), 674–688. https://doi.org/10.1249/01.mss.0000121945.36635.61

Kristman, V. L., Tator, C. H., Kreiger, N., Richards, D., Mainwaring, L., Jaglal, S., Tomlinson, G., & Comper, P. (2008). Does the apolipoprotein epsilon 4 allele predispose varsity athletes to concussion? A prospective cohort study. *Clinical Journal of Sport Medicine: Official Journal of the Canadian Academy of Sport Medicine*, *18*(4), 322–328. https://doi.org/10.1097/JSM.0b013e31817e6f3e

Kumagai, H., Miyamoto-Mikami, E., Hirata, K., Kikuchi, N., Kamiya, N., Hoshikawa, S., Zempo, H., Naito, H., Miyamoto, N., & Fuku, N. (2019). ESR1 rs2234693 polymorphism Is associated with muscle injury and muscle stiffness. *Medicine and Science in Sports and Exercise*, *51*(1), 19–26. https://doi.org/10.1249/MSS.0000000000001750

La Montagna, R., Canonico, R., Alfano, L., Bucci, E., Boffo, S., Staiano, L., Fulco, B., D'Andrea, E., De Nicola, A., Maiorano, P., D'Angelo, C., Chirico, A., De Nicola, A., & Giordano, A. (2019). Genomic analysis reveals association of specific SNPs with athletic performance and susceptibility to injuries in professional soccer players. *Journal of Cellular Physiology*. https://doi.org/10.1002/jcp.29118

Läll, K., Mägi, R., Morris, A., Metspalu, A., & Fischer, K. (2017). Personalized risk prediction for type 2 diabetes: The potential of genetic risk scores. *Genetics in Medicine: Official Journal of the American College of Medical Genetics*, *19*(3), 322–329. https://doi.org/10.1038/gim.2016.103

Larkin, P., & O'Connor, D. (2017). Talent identification and recruitment in youth soccer: Recruiter's perceptions of the key attributes for player recruitment. *PloS One*, *12*(4), e0175716. https://doi.org/10.1371/journal.pone.0175716

Larruskain, J., Celorrio, D., Barrio, I., Odriozola, A., Gil, S. M., Fernandez-Lopez, J. R., Nozal, R., Ortuzar, I., Lekue, J. A., & Aznar, J. M. (2018). Genetic variants and hamstring injury in soccer: An association and validation study. *Medicine and Science in Sports and Exercise*, *50*(2), 361–368. https://doi.org/10.1249/MSS.0000000000001434

Lath, F., Koopmann, T., Faber, I., Baker, J., & Schorer, J. (2021). Focusing on the coach's eye; towards a working model of coach decision-making in talent selection. *Psychology of Sport and Exercise*, *56*, 102011. https://doi.org/10.1016/j.psychsport.2021.102011

Lau, J., Ioannidis, J. P. A., Terrin, N., Schmid, C. H., & Olkin, I. (2006). The case of the misleading funnel plot. *BMJ (Clinical Research Ed.)*, *333*(7568), 597–600. https://doi.org/10.1136/bmj.333.7568.597

Le Gall, F., Carling, C., Williams, M., & Reilly, T. (2010). Anthropometric and fitness characteristics of international, professional and amateur male graduate soccer players from an elite youth academy. *Journal of Science and Medicine in Sport*, *13*(1), 90–95. https://doi.org/10.1016/j.jsams.2008.07.004

Lee, F. X. Z., Houweling, P. J., North, K. N., & Quinlan, K. G. R. (2016). How does α-actinin-3 deficiency alter muscle function? Mechanistic insights into ACTN3, the 'gene for speed'. *Biochimica Et Biophysica Acta*, *1863*(4), 686–693. https://doi.org/10.1016/j.bbamcr.2016.01.013

Leyhr, D., Bergmann, F., Schreiner, R., Mann, D., Dugandzic, D., & Höner, O. (2021). Relative age-related biases in objective and subjective assessments of performance in talented youth soccer players. *Frontiers in Sports and Active Living*, *3*, 664231. https://doi.org/10.3389/fspor.2021.664231

Leyhr, D., Kelava, A., Raabe, J., & Höner, O. (2018). Longitudinal motor performance development in early adolescence and its relationship to adult success: An 8-year prospective study of highly talented soccer players. *PloS One*, *13*(5), e0196324. https://doi.org/10.1371/journal.pone.0196324

Liew, G. C., Kuan, G., Chin, N. S., & Hashim, H. A. (2019). Mental toughness in sport. *German Journal of Exercise and Sport Research*, *49*(4), 381–394. https://doi.org/10.1007/s12662-019-00603-3

Lin, D.-Y. (2019). A simple and accurate method to determine genomewide significance for association tests in sequencing studies. *Genetic Epidemiology*, *43*(4), 365–372. https://doi.org/10.1002/gepi.22183

Lin, Y., Mutz, J., Clough, P. J., & Papageorgiou, K. A. (2017). Mental toughness and individual differences in learning, educational and work performance, psychological well-being, and personality: A systematic review. *Frontiers in Psychology*, *8*, 1345. https://doi.org/10.3389/fpsyg.2017.01345

Little, J., Higgins, J. P. T., Ioannidis, J. P. A., Moher, D., Gagnon, F., von Elm, E., Khoury, M. J., Cohen, B., Davey-Smith, G., Grimshaw, J., Scheet, P., Gwinn, M., Williamson, R. E., Zou, G. Y., Hutchings, K., Johnson, C. Y., Tait, V., Wiens, M., Golding, J., … Birkett, N. (2009). STrengthening the REporting of Genetic Association studies (STREGA)—An extension of the STROBE statement. *European Journal of Clinical Investigation*, *39*(4), 247–266. https://doi.org/10.1111/j.1365-2362.2009.02125.x

Little, T., & Williams, A. G. (2005). Specificity of acceleration, maximum speed, and agility in professional soccer players. *Journal of Strength and Conditioning Research*, *19*(1), 76–78. https://doi.org/10.1519/14253.1

Livshits, G., Gao, F., Malkin, I., Needhamsen, M., Xia, Y., Yuan, W., Bell, C. G., Ward, K., Liu, Y., Wang, J., Bell, J. T., & Spector, T. D. (2016). Contribution of heritability and epigenetic factors to skeletal muscle mass variation in United Kingdom twins. *The Journal of Clinical Endocrinology and Metabolism*, *101*(6), 2450–2459. https://doi.org/10.1210/jc.2016-1219

Locke, A. E., Kahali, B., Berndt, S. I., Justice, A. E., Pers, T. H., Day, F. R., Powell, C., Vedantam, S., Buchkovich, M. L., Yang, J., Croteau-Chonka, D. C., Esko, T., Fall, T., Ferreira, T., Gustafsson, S., Kutalik, Z., Luan, J., Mägi, R., Randall, J. C., … Speliotes, E. K. (2015). Genetic studies of body mass index yield new insights for obesity biology. *Nature*, *518*(7538), 197–206. https://doi.org/10.1038/nature14177

Longo, U. G., Candela, V., Berton, A., Salvatore, G., Guarnieri, A., DeAngelis, J., Nazarian, A., & Denaro, V. (2019). Genetic basis of rotator cuff injury: A systematic review. *BMC Medical Genetics*, *20*(1), 149. https://doi.org/10.1186/s12881-019-0883-y

López-Valenciano, A., Ruiz-Pérez, I., Garcia-Gómez, A., Vera-Garcia, F. J., Croix, M. D. S., Myer, G. D., & Ayala, F. (2020). Epidemiology of injuries in professional football: A systematic review and meta-analysis. *British Journal of Sports Medicine*, *54*(12), 711–718. https://doi.org/10.1136/bjsports-2018-099577

Lucia, A., Morán, M., Zi-hong, H., & Ruiz, J. R. (2010). Elite athletes: Are the genes the champions? *International Journal of Sports Physiology and Performance*, *5*(1), 98–102. https://doi.org/10.1123/ijspp.5.1.93

Lulińska-Kuklik, E., Laguette, M.-J. N., Moska, W., Weber-Rajek, M., Ficek, K., Puchala, R., Cięszczyk, P., Sawczuk, M., September, A. V., & Maciejewska-Skrendo, A. (2019). Are TNC gene variants associated with anterior cruciate ligament rupture susceptibility? *Journal of Science and Medicine in Sport*, *22*(4), 408–412. https://doi.org/10.1016/j.jsams.2018.10.003

Lulińska-Kuklik, E., Leźnicka, K., Humińska-Lisowska, K., Moska, W., Michałowska-Sawczyn, M., Ossowski, Z., Maculewicz, E., Cięszczyk, P., Kaczmarczyk, M., Ratkowski, W., Ficek, K., Zmijewski, P., & Leońska-Duniec, A. (2019). The VEGFA gene and anterior cruciate ligament rupture risk in the Caucasian population. *Biology of Sport*, *36*(1), 3–8. https://doi.org/10.5114/biolsport.2018.78902

Lulińska-Kuklik, E., Rahim, M., Domańska-Senderowska, D., Ficek, K., Michałowska-Sawczyn, M., Moska, W., Kaczmarczyk, M., Brzeziański, M., Brzeziańska-Lasota, E., Cięszczyk, P., & September, A. V. (2018). Interactions between COL5A1 gene and risk

of the anterior cruciate ligament rupture. *Journal of Human Kinetics*, *62*, 65–71. https://doi.org/10.1515/hukin-2017-0177

Lv, Z.-T., Gao, S.-T., Cheng, P., Liang, S., Yu, S.-Y., Yang, Q., & Chen, A.-M. (2018). Association between polymorphism rs12722 in COL5A1 and musculoskeletal soft tissue injuries: A systematic review and meta-analysis. *Oncotarget*, *9*(20), 15365–15374. https://doi.org/10.18632/oncotarget.23805

Lyall, D. M., Lopez, L. M., Bastin, M. E., Maniega, S. M., Penke, L., Valdés Hernández, M. del C., Royle, N. A., Starr, J. M., Porteous, D. J., Wardlaw, J. M., & Deary, I. J. (2013). ADRB2, brain white matter integrity and cognitive ageing in the Lothian Birth Cohort 1936. *Behavior Genetics*, *43*(1), 13–23. https://doi.org/10.1007/s10519-012-9570-x

Ma, F., Yang, Y., Li, X., Zhou, F., Gao, C., Li, M., & Gao, L. (2013). The association of sport performance with ACE and ACTN3 genetic polymorphisms: A systematic review and meta-analysis. *PloS One*, *8*(1), e54685. https://doi.org/10.1371/journal.pone.0054685

Machiela, M. J., & Chanock, S. J. (2015). LDlink: A web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics (Oxford, England)*, *31*(21), 3555–3557. https://doi.org/10.1093/bioinformatics/btv402

Maciejewska-Skrendo, A., Cięszczyk, P., Chycki, J., Sawczuk, M., & Smółka, W. (2019). Genetic markers associated with power athlete status. *Journal of Human Kinetics*, *68*, 17–36. https://doi.org/10.2478/hukin-2019-0053

Macnamara, B. N., Moreau, D., & Hambrick, D. Z. (2016). The relationship between deliberate practice and performance in sports: A meta-analysis. *Perspectives on Psychological*

Science: A Journal of the Association for Psychological Science, *11*(3), 333–350. https://doi.org/10.1177/1745691616635591

Magnusson, K., Turkiewicz, A., Hughes, V., Frobell, R., & Englund, M. (2021). High genetic contribution to anterior cruciate ligament rupture: Heritability ~69%. *British Journal of Sports Medicine*, *55*(7), 385–389. https://doi.org/10.1136/bjsports-2020-102392

Mallo, J., Mena, E., Nevado, F., & Paredes, V. (2015). Physical demands of top-class soccer friendly matches in relation to a playing position using global positioning system technology. *Journal of Human Kinetics*, *47*, 179–188. https://doi.org/10.1515/hukin-2015-0073

Mancini, A., Vitucci, D., Labruna, G., Imperlini, E., Randers, M. B., Schmidt, J. F., Hagman, M., Andersen, T. R., Russo, R., Orrù, S., Krustrup, P., Salvatore, F., & Buono, P. (2017). Effect of lifelong football training on the expression of muscle molecular markers involved in healthy longevity. *European Journal of Applied Physiology*, *117*(4), 721–730. https://doi.org/10.1007/s00421-017-3562-8

Mannion, S., Mtintsilana, A., Posthumus, M., van der Merwe, W., Hobbs, H., Collins, M., & September, A. V. (2014). Genes encoding proteoglycans are associated with the risk of anterior cruciate ligament ruptures. *British Journal of Sports Medicine*, *48*(22), 1640–1646. https://doi.org/10.1136/bjsports-2013-093201

Martorell-Marugan, J., Toro-Dominguez, D., Alarcon-Riquelme, M. E., & Carmona-Saez, P. (2017). MetaGenyo: A web tool for meta-analysis of genetic association studies. *BMC Bioinformatics*, *18*. https://doi.org/10.1186/s12859-017-1990-4

Massidda, M., Bachis, V., Corrias, L., Piras, F., Scorcu, M., Culigioni, C., Masala, D., & Calò, C. M. (2015). ACTN3 R577X polymorphism is not associated with team sport athletic

status in Italians. *Sports Medicine - Open*, *1*(1), 6. https://doi.org/10.1186/s40798-015-0008-x

Massidda, M., Corrias, L., Bachis, V., Cugia, P., Piras, F., Scorcu, M., & Calò, C. M. (2015). Vitamin D receptor gene polymorphisms and musculoskeletal injuries in professional football players. *Experimental and Therapeutic Medicine*, *9*(5), 1974–1978. https://doi.org/10.3892/etm.2015.2364

Massidda, M., Corrias, L., Bachis, V., Piras, F., Cugia, P., Scorcu, M., & Calò, C. M. (2014). *Genetic polymorphisms and muscle injuries among Italian soccer players*. 6.

Massidda, M., Eynon, N., Bachis, V., Corrias, L., Culigioni, C., Cugia, P., Scorcu, M., & Calò, C. M. (2016). Association between MCT1 A1470T polymorphism and fat-free mass in well-trained young soccer players. *Journal of Strength and Conditioning Research*, *30*(4), 1171–1176. https://doi.org/10.1519/JSC.0000000000001176

Massidda, M., Eynon, N., Bachis, V., Corrias, L., Culigioni, C., Piras, F., Cugia, P., Scorcu, M., & Calò, C. M. (2015). Influence of the MCT1 rs1049434 on indirect muscle disorders/injuries in elite football players. *Sports Medicine - Open*, *1*(1), 33. https://doi.org/10.1186/s40798-015-0033-9

Massidda, M., Mendez-Villanueva, A., Ginevičienė, V., Proia, P., Drozdovska, S. B., Dosenko, V., Scorcu, M., Stula, A., Sawczuk, M., Cięszczyk, P., & Calò, C. M. (2018). Association of monocarboxylate transporter-1 (MCT1) A1470T polymorphism (rs1049434) with forward football player status. *International Journal of Sports Medicine*, *39*(13), 1028–1034. https://doi.org/10.1055/a-0634-6387

Massidda, M., Miyamoto-Mikami, E., Kumagai, H., Ikeda, H., Shimasaki, Y., Yoshimura, M., Cugia, P., Piras, F., Scorcu, M., Kikuchi, N., Calò, C. M., & Fuku, N. (2020). Association

between the ACE I/D polymorphism and muscle injuries in Italian and Japanese elite football players. *Journal of Sports Sciences*, *38*(21), 2423–2429. https://doi.org/10.1080/02640414.2020.1787683

Massidda, M., Scorcu, M., & Calò, C. M. (2014). New genetic model for predicting phenotype traits in sports. *International Journal of Sports Physiology and Performance*, *9*(3), 554–560. https://doi.org/10.1123/ijspp.2012-0339

Massidda, M., Voisin, S., Culigioni, C., Piras, F., Cugia, P., Yan, X., Eynon, N., & Calò, C. M. (2019). ACTN3 R577X polymorphism is associated with the incidence and severity of injuries in professional football players. *Clinical Journal of Sport Medicine: Official Journal of the Canadian Academy of Sport Medicine*, *29*(1), 57–61. https://doi.org/10.1097/JSM.0000000000000487

Massidda, M., Vona, G., & Calò, C. M. (2009). Association between the ACTN3 R577X polymorphism and artistic gymnastic performance in Italy. *Genetic Testing and Molecular Biomarkers*, *13*(3), 377–380. https://doi.org/10.1089/gtmb.2008.0157

Massidda, M., Vona, G., & Calò, C. M. (2011). Lack of association between ACE gene insertion/deletion polymorphism and elite artistic gymnastic performance of Italian gymnasts. *European Journal of Sport Science*, *11*(3), 149–153. https://doi.org/10.1080/17461391.2010.499971

Matsumoto, N., Hanakawa, T., Maki, S., Graybiel, A. M., & Kimura, M. (1999). Role of [corrected] nigrostriatal dopamine system in learning to perform sequential motor tasks in a predictive manner. *Journal of Neurophysiology*, *82*(2), 978–998. https://doi.org/10.1152/jn.1999.82.2.978

Mattsson, C. M., Wheeler, M. T., Waggott, D., Caleshu, C., & Ashley, E. A. (2016). Sports genetics moving forward: Lessons learned from medical research. *Physiological Genomics*, *48*(3), 175–182. https://doi.org/10.1152/physiolgenomics.00109.2015

Maxcy, S. J. (2003). Pragmatic threads in mixed methods research in the social sciences: The search for multiple modes of inquiry and the end of the philosophy of formalism. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of Mixed Methods in Social and Behavioral Research* (pp. 51–89). Sage Publications.

McAuley, A. B. T., Baker, J., & Kelly, A. L. (2021). How nature and nurture conspire to influence athletic success. In A. L. Kelly, J. Côté, M. Jeffreys, & J. Turnnidge (Eds.), *Birth advantages and relative age effects in sport: Exploring organizational structures and creating appropriate settings* (pp. 159–183). Routledge.

McAuley, A. B. T., Baker, J., & Kelly, A. L. (2022). Defining "elite" status in sport: From chaos to clarity. *German Journal of Exercise and Sport Research*, *52*(1), 193–197. https://doi.org/10.1007/s12662-021-00737-3

McAuley, A. B. T., Hughes, D. C., Tsaprouni, L. G., Varley, I., Suraci, B., Baker, J., Herbert, A. J., & Kelly, A. L. (2022a). Genetic associations with technical capabilities in English academy football players: A preliminary study. *The Journal of Sports Medicine and Physical Fitness*. https://doi.org/10.23736/S0022-4707.22.13945-9

McAuley, A. B. T., Hughes, D. C., Tsaprouni, L. G., Varley, I., Suraci, B., Baker, J., Herbert, A. J., & Kelly, A. L. (2022b). Genetic associations with personality and mental toughness profiles of English academy football players: An exploratory study. *Psychology of Sport and Exercise*, *61*, 102209. https://doi.org/10.1016/j.psychsport.2022.102209

McAuley, A. B. T., Hughes, D. C., Tsaprouni, L. G., Varley, I., Suraci, B., Baker, J., Herbert, A. J., & Kelly, A. L. (2022c). Genetic variations between youth and professional development phase English academy football players. *Genes*, *13*(11), 2001. https://doi.org/10.3390/genes13112001

McAuley, A. B. T., Hughes, D. C., Tsaprouni, L. G., Varley, I., Suraci, B., Roos, T. R., Herbert, A. J., Jackson, D. T., & Kelly, A. L. (2022d). A systematic review of the genetic predisposition to injury in football. *Journal of Science in Sport and Exercise*. https://doi.org/10.1007/s42978-022-00187-9

McAuley, A. B. T., Hughes, D. C., Tsaprouni, L. G., Varley, I., Suraci, B., Roos, T. R., Herbert, A. J., & Kelly, A. L. (2021a). Genetic association research in football: A systematic review. *European Journal of Sport Science*, *21*(5), 714–752. https://doi.org/10.1080/17461391.2020.1776401

McAuley, A. B. T., Hughes, D. C., Tsaprouni, L. G., Varley, I., Suraci, B., Roos, T. R., Herbert, A. J., & Kelly, A. L. (2021b). The association of the ACTN3 R577X and ACE I/D polymorphisms with athlete status in football: A systematic review and meta-analysis. *Journal of Sports Sciences*, *39*(2), 200–211. https://doi.org/10.1080/02640414.2020.1812195

McAuley, A. B. T., Hughes, D. C., Tsaprouni, L. G., Varley, I., Suraci, B., Roos, T. R., Herbert, A. J., & Kelly, A. L. (2022e). Genetic testing in professional football: Perspectives of key stakeholders. *Journal of Science in Sport and Exercise*, *4*(1), 49–59. https://doi.org/10.1007/s42978-021-00131-3

McCabe, K., & Collins, C. (2018). Can genetics predict sports injury? The association of the genes GDF5, AMPD1, COL5A1 and IGF2 on soccer player injury occurrence. *Sports (Basel, Switzerland)*, *6*(1), 21. https://doi.org/10.3390/sports6010021

McCall, A., Carling, C., Davison, M., Nedelec, M., Le Gall, F., Berthoin, S., & Dupont, G. (2015). Injury risk factors, screening tests and preventative strategies: A systematic review of the evidence that underpins the perceptions and practices of 44 football (soccer) teams from various premier leagues. *British Journal of Sports Medicine*, *49*(9), 583–589. https://doi.org/10.1136/bjsports-2014-094104

McCarthy, M. I., Abecasis, G. R., Cardon, L. R., Goldstein, D. B., Little, J., Ioannidis, J. P. A., & Hirschhorn, J. N. (2008). Genome-wide association studies for complex traits: Consensus, uncertainty and challenges. *Nature Reviews. Genetics*, *9*(5), 356–369. https://doi.org/10.1038/nrg2344

McCrae, R. R., & John, O. P. (1992). An introduction to the five-factor model and its applications. *Journal of Personality*, *60*(2), 175–215. https://doi.org/10.1111/j.1467-6494.1992.tb00970.x

McCrea, M., Meier, T., Huber, D., Ptito, A., Bigler, E., Debert, C. T., Manley, G., Menon, D., Chen, J.-K., Wall, R., Schneider, K. J., & McAllister, T. (2017). Role of advanced neuroimaging, fluid biomarkers and genetic testing in the assessment of sport-related concussion: A systematic review. *British Journal of Sports Medicine*, *51*(12), 919–929. https://doi.org/10.1136/bjsports-2016-097447

McDevitt, J. K., Tierney, R. T., Mansell, J. L., Driban, J. B., Higgins, M., Toone, N., Mishra, A., & Krynetskiy, E. (2011). Neuronal structural protein polymorphism and concussion

in college athletes. *Brain Injury*, *25*(11), 1108–1113. https://doi.org/10.3109/02699052.2011.607790

Meckel, Y., Eliakim, A., Nemet, D., Levin, N., & Ben-Zaken, S. (2019). PPARD CC and ACTN3 RR genotype prevalence among elite soccer players. *Science and Medicine in Football*, *0*(0), 1–6. https://doi.org/10.1080/24733938.2019.1677936

Meintzschel, F., & Ziemann, U. (2006). Modification of practice-dependent plasticity in human motor cortex by neuromodulators. *Cerebral Cortex (New York, N.Y.: 1991)*, *16*(8), 1106–1115. https://doi.org/10.1093/cercor/bhj052

Menezes, W. C. F., Gingozac, T. C. V., Oliveira, L. A. S., Pedrino, G. R., Ternes, Y. M. F., Santos, R. S., & Reis, A. A. S. (2019). Research article assessment of effects of the genetic polymorphisms on soccer athletes. *Genetics and Molecular Research*, *18*(2). https://doi.org/10.4238/gmr18192

Micheli, M. L., Gulisano, M., Morucci, G., Punzi, T., Ruggiero, M., Ceroti, M., Marella, M., Castellini, E., & Pacini, S. (2011). Angiotensin-converting enzyme/vitamin D receptor gene polymorphisms and bioelectrical impedance analysis in predicting athletic performances of Italian young soccer players. *Journal of Strength and Conditioning Research*, *25*(8), 2084–2091. https://doi.org/10.1519/JSC.0b013e31820238aa

Mikami, E., Fuku, N., Murakami, H., Tsuchie, H., Takahashi, H., Ohiwa, N., Tanaka, H., Pitsiladis, Y. P., Higuchi, M., Miyachi, M., Kawahara, T., & Tanaka, M. (2014). ACTN3 R577X genotype is associated with sprinting in elite Japanese athletes. *International Journal of Sports Medicine*, *35*(2), 172–177. https://doi.org/10.1055/s-0033-1347171

Miller, A. (2016). *Barcelona breaking the mould with DNA testing as La Liga giants prepare for Champions League clash with Arsenal | Daily Mail Online*.

https://www.dailymail.co.uk/sport/football/article-3456465/Barcelona-breaking-mould-DNA-testing-La-Liga-giants-prepare-Champions-League-clash-Arsenal.html

Mills, M., Yang, N., Weinberger, R., Vander Woude, D. L., Beggs, A. H., Easteal, S., & North, K. (2001). Differential expression of the actin-binding proteins, alpha-actinin-2 and -3, in different species: Implications for the evolution of functional redundancy. *Human Molecular Genetics*, *10*(13), 1335–1346. https://doi.org/10.1093/hmg/10.13.1335

Min, S.-K., Takahashi, K., Ishigami, H., Hiranuma, K., Mizuno, M., Ishii, T., Kim, C.-S., & Nakazato, K. (2009). Is there a gender difference between ACE gene and race distance? *Applied Physiology, Nutrition, and Metabolism = Physiologie Appliquee, Nutrition Et Metabolisme*, *34*(5), 926–932. https://doi.org/10.1139/H09-097

Mirkov, D. M., Kukolj, M., Ugarkovic, D., Koprivica, V. J., & Jaric, S. (2010). Development of anthropometric and physical performance profiles of young elite male soccer players: A longitudinal study. *Journal of Strength and Conditioning Research*, *24*(10), 2677–2682. https://doi.org/10.1519/JSC.0b013e3181e27245

Missitzi, J., Geladas, N., & Klissouras, V. (2004). Heritability in neuromuscular coordination: Implications for motor control strategies. *Medicine and Science in Sports and Exercise*, *36*(2), 233–240. https://doi.org/10.1249/01.MSS.0000113479.98631.C4

Missitzi, J., Geladas, N., Misitzi, A., Misitzis, L., Classen, J., & Klissouras, V. (2018). Heritability of proprioceptive senses. *Journal of Applied Physiology*, *125*(4), 972–982. https://doi.org/10.1152/japplphysiol.00544.2017

Missitzi, J., Gentner, R., Geladas, N., Politis, P., Karandreas, N., Classen, J., & Klissouras, V. (2011). Plasticity in human motor cortex is in part genetically determined. *The Journal of Physiology*, *589*(Pt 2), 297–306. https://doi.org/10.1113/jphysiol.2010.200600

Missitzi, J., Gentner, R., Misitzi, A., Geladas, N., Politis, P., Klissouras, V., & Classen, J. (2013). Heritability of motor control and motor learning. *Physiological Reports*, *1*(7), e00188. https://doi.org/10.1002/phy2.188

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine*, *6*(7), e1000097–e1000097. https://doi.org/10.1371/journal.pmed.1000097

Mohr, M., Krustrup, P., & Bangsbo, J. (2003). Match performance of high-standard soccer players with special reference to development of fatigue. *Journal of Sports Sciences*, *21*(7), 519–528. https://doi.org/10.1080/0264041031000071182

Monnerat, G., Maior, A. S., Tannure, M., Back, L. K. F. C., & Santos, C. G. M. (2019). Single-nucleotide-polymorphism-panel population-genetics approach based on the 1000 genomes database and elite soccer players. *International Journal of Sports Physiology and Performance*, *14*(6), 711–717. https://doi.org/10.1123/ijspp.2018-0715

Monnerat-Cahli, G., Paulúcio, D., Moura Neto, R. S., Silva, R., Pompeu, F., Budowle, B., & Santos, C. G. (2017). Letter to the editor: Are the doors opened to a genetic-based algorithm for personalized resistance training? *Biology of Sport*, *34*(1), 27–29. https://doi.org/10.5114/biolsport.2017.63384

Montgomery, H. E., Marshall, R., Hemingway, H., Myerson, S., Clarkson, P., Dollery, C., Hayward, M., Holliman, D. E., Jubb, M., World, M., Thomas, E. L., Brynes, A. E., Saeed, N., Barnard, M., Bell, J. D., Prasad, K., Rayson, M., Talmud, P. J., & Humphries, S. E. (1998). Human gene for physical performance. *Nature*, *393*(6682), 221–222. https://doi.org/10.1038/30374

Moreland, E., Borisov, O. V., Semenova, E. A., Larin, A. K., Andryushchenko, O. N., Andryushchenko, L. B., Generozov, E. V., Williams, A. G., & Ahmetov, I. I. (2020). Polygenic profile of elite strength athletes. *Journal of Strength and Conditioning Research*. https://doi.org/10.1519/JSC.0000000000003901

Moreno, V., Areces, F., Ruiz-Vicente, D., Ordovás, J. M., & Del Coso, J. (2020). Influence of the ACTN3 R577X genotype on the injury epidemiology of marathon runners. *PLoS ONE*, *15*(1). https://doi.org/10.1371/journal.pone.0227548

Morgan, D. L. (2007). Paradigms lost and pragmatism regained: Methodological implications of combining qualitative and quantitative methods. *Journal of Mixed Methods Research*, *1*(1), 48–76. https://doi.org/10.1177/2345678906292462

Morgan, D. L. (2014). Pragmatism as a paradigm for social research. *Qualitative Inquiry*, *20*(8), 1045–1053. https://doi.org/10.1177/1077800413513733

Morin-Moncet, O., Beaumont, V., de Beaumont, L., Lepage, J.-F., & Théoret, H. (2014). BDNF Val66Met polymorphism is associated with abnormal interhemispheric transfer of a newly acquired motor skill. *Journal of Neurophysiology*, *111*(10), 2094–2102. https://doi.org/10.1152/jn.00388.2013

Munzenmaier, D. H., & Greene, A. S. (1996). Opposing actions of angiotensin II on microvascular growth and arterial blood pressure. *Hypertension (Dallas, Tex.: 1979)*, *27*(3 Pt 2), 760–765. https://doi.org/10.1161/01.hyp.27.3.760

Murr, D., Feichtinger, P., Larkin, P., O'Connor, D., & Höner, O. (2018). Psychological talent predictors in youth soccer: A systematic review of the prognostic relevance of psychomotor, perceptual-cognitive and personality-related factors. *PloS One*, *13*(10), e0205337. https://doi.org/10.1371/journal.pone.0205337

Murr, D., Raabe, J., & Höner, O. (2018). The prognostic value of physiological and physical characteristics in youth soccer: A systematic review. *European Journal of Sport Science*, *18*(1), 62–74. https://doi.org/10.1080/17461391.2017.1386719

Murtagh, C. F., Brownlee, T. E., O'Boyle, A., Morgans, R., Drust, B., & Erskine, R. M. (2018). Importance of speed and power in elite youth soccer depends on maturation status. *Journal of Strength and Conditioning Research*, *32*(2), 297–303. https://doi.org/10.1519/JSC.0000000000002367

Murtagh, C. F., Brownlee, T. E., Rienzi, E., Roquero, S., Moreno, S., Huertas, G., Lugioratto, G., Baumert, P., Turner, D. C., Lee, D., Dickinson, P., Lyon, K. A., Sheikhsaraf, B., Biyik, B., O'Boyle, A., Morgans, R., Massey, A., Drust, B., & Erskine, R. M. (2020). The genetic profile of elite youth soccer players and its association with power and speed depends on maturity status. *PloS One*, *15*(6), e0234458. https://doi.org/10.1371/journal.pone.0234458

Musculus, L., & Lobinger, B. H. (2018). Psychological characteristics in talented soccer players—Recommendations on how to improve coaches' assessment. *Frontiers in Psychology*, *9*, 41. https://doi.org/10.3389/fpsyg.2018.00041

Mustapic, M., Presecki, P., Pivac, N., Mimica, N., Hof, P. R., Simic, G., Folnegovic-Smalc, V., & Muck-Seler, D. (2013). Genotype-independent decrease in plasma dopamine beta-hydroxylase activity in Alzheimer's disease. *Progress in Neuro-Psychopharmacology & Biological Psychiatry*, *44*, 94–99. https://doi.org/10.1016/j.pnpbp.2013.02.002

Myerson, S., Hemingway, H., Budget, R., Martin, J., Humphries, S., & Montgomery, H. (1999). Human angiotensin I-converting enzyme gene and endurance performance.

Journal of Applied Physiology (Bethesda, Md.: 1985), *87*(4), 1313–1316. https://doi.org/10.1152/jappl.1999.87.4.1313

Nauright, J. (2004). Global games: Culture, political economy and sport in the globalised world of the 21st century. *Third World Quarterly*, *25*(7), 1325–1336. https://doi.org/10.1080/0143659042002813O2

Nazarov, I. B., Woods, D. R., Montgomery, H. E., Shneider, O. V., Kazakov, V. I., Tomilin, N. V., & Rogozkin, V. A. (2001). The angiotensin converting enzyme I/D polymorphism in Russian athletes. *European Journal of Human Genetics: EJHG*, *9*(10), 797–801. https://doi.org/10.1038/sj.ejhg.5200711

Niemi, A.-K., & Majamaa, K. (2005). Mitochondrial DNA and ACTN3 genotypes in Finnish elite endurance and sprint athletes. *European Journal of Human Genetics: EJHG*, *13*(8), 965–969. https://doi.org/10.1038/sj.ejhg.5201438

North, K. N., Yang, N., Wattanasirichaigoon, D., Mills, M., Easteal, S., & Beggs, A. H. (1999). A common nonsense mutation results in alpha-actinin-3 deficiency in the general population. *Nature Genetics*, *21*(4), 353–354. https://doi.org/10.1038/7675

Oh, S. D. (2007). The distribution of I/D polymorphism in the ACE gene among Korean male elite athletes. *The Journal of Sports Medicine and Physical Fitness*, *47*(2), 250–254.

Oliver, J. L., Armstrong, N., & Williams, C. A. (2009). Relationship between brief and prolonged repeated sprint ability. *Journal of Science and Medicine in Sport*, *12*(1), 238–243. https://doi.org/10.1016/j.jsams.2007.09.006

Orysiak, J., Busko, K., Mazur-RóŻycka, J., Michalski, R., Gajewski, J., Malczewska-Lenczowska, J., & Sitkowski, D. (2015). Relationship between ACTN3 R577X polymorphism and physical abilities in polish athletes. *Journal of Strength and*

Conditioning Research, *29*(8), 2333–2339. https://doi.org/10.1519/JSC.0000000000000880

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *Journal of Clinical Epidemiology*, *134*, 178–189. https://doi.org/10.1016/j.jclinepi.2021.03.001

Panenka, W. J., Gardner, A. J., Dretsch, M. N., Crynen, G. C., Crawford, F. C., & Iverson, G. L. (2017). Systematic review of genetic risk factors for sustaining a mild traumatic brain injury. *Journal of Neurotrauma*, *34*(13), 2093–2099. https://doi.org/10.1089/neu.2016.4833

Pansiri, J. (2005). Pragmatism: A methodological approach to researching strategic alliances in tourism. *Tourism and Hospitality Planning & Development*, *2*(3), 191–206. https://doi.org/10.1080/14790530500399333

Papadimitriou, I. D., Papadopoulos, C., Kouvatsi, A., & Triantaphyllidis, C. (2008). The ACTN3 gene in elite Greek track and field athletes. *International Journal of Sports Medicine*, *29*(4), 352–355. https://doi.org/10.1055/s-2007-965339

Paparini, A., Ripani, M., Giordano, G. D., Santoni, D., Pigozzi, F., & Romano-Spica, V. (2007). ACTN3 genotyping by real-time PCR in the Italian population and athletes. *Medicine and Science in Sports and Exercise*, *39*(5), 810–815. https://doi.org/10.1097/mss.0b013e3180317491

Patel, R., Nevill, A., Smith, T., Cloak, R., & Wyon, M. (2020). The influence of birth quartile, maturation, anthropometry and physical performances on player retention: Observations from an elite football academy. *International Journal of Sports Science & Coaching*, *15*(2), 121–134. https://doi.org/10.1177/1747954120906507

Patel, S., & Varley, I. (2019). Exploring the regulation of genetic testing in sport. *The Entertainment and Sports Law Journal*, *17*(1), Article 1. https://doi.org/10.16997/eslj.223

Pearson-Fuhrhop, K. M., Minton, B., Acevedo, D., Shahbaba, B., & Cramer, S. C. (2013). Genetic variation in the human brain dopamine system influences motor learning and its modulation by L-Dopa. *PloS One*, *8*(4), e61197. https://doi.org/10.1371/journal.pone.0061197

Peeters, M. W., Thomis, M. A., Loos, R. J. F., Derom, C. A., Fagard, R., Claessens, A. L., Vlietinck, R. F., & Beunen, G. P. (2007). Heritability of somatotype components: A multivariate analysis. *International Journal of Obesity*, *31*(8), 1295–1301. https://doi.org/10.1038/sj.ijo.0803575

Pellicciari, M. C., Veniero, D., Marzano, C., Moroni, F., Pirulli, C., Curcio, G., Ferrara, M., Miniussi, C., Rossini, P. M., & De Gennaro, L. (2009). Heritability of intracortical inhibition and facilitation. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *29*(28), 8897–8900. https://doi.org/10.1523/JNEUROSCI.2112-09.2009

Petersen, A. M. W., & Pedersen, B. K. (2005). The anti-inflammatory effect of exercise. *Journal of Applied Physiology (Bethesda, Md.: 1985)*, *98*(4), 1154–1162. https://doi.org/10.1152/japplphysiol.00164.2004

Petito, A., Altamura, M., Iuso, S., Padalino, F. A., Sessa, F., D'Andrea, G., Margaglione, M., & Bellomo, A. (2016). The relationship between personality traits, the 5HTT polymorphisms, and the occurrence of anxiety and depressive symptoms in elite athletes. *PloS One*, *11*(6), e0156601. https://doi.org/10.1371/journal.pone.0156601

Petr, M., Thiel, D., Kateřina, K., Brož, P., Malý, T., Zahálka, F., Vostatková, P., Wilk, M., Chycki, J., & Stastny, P. (2021). Speed and power-related gene polymorphisms associated with playing position in elite soccer players. *Biology of Sport*, *39*(2), 355–366. https://doi.org/10.5114/biolsport.2022.105333

Pickering, C., & Kiely, J. (2018). Hamstring injury prevention: A role for genetic information? *Medical Hypotheses*, *119*, 58–62. https://doi.org/10.1016/j.mehy.2018.07.011

Pickering, C., & Kiely, J. (2021). The frequency of, and attitudes towards, genetic testing amongst athletes and support staff. *Performance Enhancement & Health*, *8*(4), 100184. https://doi.org/10.1016/j.peh.2020.100184

Pickering, C., Kiely, J., Grgic, J., Lucia, A., & Del Coso, J. (2019). Can genetic testing identify talent for sport? *Genes*, *10*(12). https://doi.org/10.3390/genes10120972

Pickering, C., Kiely, J., Suraci, B., & Collins, D. (2018). The magnitude of Yo-Yo test improvements following an aerobic training intervention are associated with total genotype score. *PloS One*, *13*(11), e0207597. https://doi.org/10.1371/journal.pone.0207597

Pickering, C., Kiely, J., Suraci, B., Quigley, C., & Watson, J. (2017). A genetic-based algorithm for recovery: A pilot study. *BioRxiv*, 166181. https://doi.org/10.1101/166181

Pickering, C., Suraci, B., Semenova, E. A., Boulygina, E. A., Kostryukova, E. S., Kulemin, N. A., Borisov, O. V., Khabibova, S. A., Larin, A. K., Pavlenko, A. V., Lyubaeva, E. V.,

Popov, D. V., Lysenko, E. A., Vepkhvadze, T. F., Lednev, E. M., Leońska-Duniec, A., Pająk, B., Chycki, J., Moska, W., … Ahmetov, I. I. (2019). A genome-wide association study of sprint performance in elite youth football players. *Journal of Strength and Conditioning Research*, *33*(9), 2344–2351. https://doi.org/10.1519/JSC.0000000000003259

Piepiora, P. (2021). Assessment of personality traits influencing the performance of men in team sports in terms of the Big Five. *Frontiers in Psychology*, *12*, 679724. https://doi.org/10.3389/fpsyg.2021.679724

Pimenta, E. M., Coelho, D. B., Cruz, I. R., Morandi, R. F., Veneroso, C. E., de Azambuja Pussieldi, G., Carvalho, M. R. S., Silami-Garcia, E., & De Paz Fernández, J. A. (2012). The ACTN3 genotype in soccer players in response to acute eccentric training. *European Journal of Applied Physiology*, *112*(4), 1495–1503. https://doi.org/10.1007/s00421-011-2109-7

Pimenta, E. M., Coelho, D. B., Veneroso, C. E., Barros Coelho, E. J., Cruz, I. R., Morandi, R. F., De A Pussieldi, G., Carvalho, M. R. S., Garcia, E. S., & De Paz Fernández, J. A. (2013). Effect of ACTN3 gene on strength and endurance in soccer players. *Journal of Strength and Conditioning Research*, *27*(12), 3286–3292. https://doi.org/10.1519/JSC.0b013e3182915e66

Pitsiladis, Y. P., Tanaka, M., Eynon, N., Bouchard, C., North, K. N., Williams, A. G., Collins, M., Moran, C. N., Britton, S. L., Fuku, N., Ashley, E. A., Klissouras, V., Lucia, A., Ahmetov, I. I., de Geus, E., Alsayrafi, M., & Athlome Project Consortium. (2016). Athlome Project Consortium: A concerted effort to discover genomic and other 'omic'

markers of athletic performance. *Physiological Genomics*, *48*(3), 183–190. https://doi.org/10.1152/physiolgenomics.00105.2015

Polderman, T. J. C., Benyamin, B., de Leeuw, C. A., Sullivan, P. F., van Bochoven, A., Visscher, P. M., & Posthuma, D. (2015). Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nature Genetics*, *47*(7), Article 7. https://doi.org/10.1038/ng.3285

Pons, E., Ponce-Bordón, J. C., Díaz-García, J., López Del Campo, R., Resta, R., Peirau, X., & García-Calvo, T. (2021). A longitudinal exploration of match running performance during a football match in the Spanish La Liga: A four-season study. *International Journal of Environmental Research and Public Health*, *18*(3), 1133. https://doi.org/10.3390/ijerph18031133

Power, R. A., & Pluess, M. (2015). Heritability estimates of the Big Five personality traits based on common genetic variants. *Translational Psychiatry*, *5*(7), e604. https://doi.org/10.1038/tp.2015.96

Premier League. (2013). *Elite Player Performance Plan*. https://www.premierleague.com/youth/EPPP

Proia, P., Bianco, A., Schiera, G., Saladino, P., Contrò, V., Caramazza, G., Traina, M., Grimaldi, K. A., Palma, A., & Paoli, A. (2014). PPARα gene variants as predicted performance-enhancing polymorphisms in professional Italian soccer players. *Open Access Journal of Sports Medicine*, *5*, 273–278. https://doi.org/10.2147/OAJSM.S68333

Pruna, R., Artells, R., Lundblad, M., & Maffulli, N. (2017). Genetic biomarkers in non-contact muscle injuries in elite soccer players. *Knee Surgery, Sports Traumatology, Arthroscopy:*

Official Journal of the ESSKA, *25*(10), 3311–3318. https://doi.org/10.1007/s00167-016-4081-6

Pruna, R., Artells, R., Ribas, J., Montoro, B., Cos, F., Muñoz, C., Rodas, G., & Maffulli, N. (2013). Single nucleotide polymorphisms associated with non-contact soft tissue injuries in elite professional soccer players: Influence on degree of injury and recovery time. *BMC Musculoskeletal Disorders*, *14*, 221. https://doi.org/10.1186/1471-2474-14-221

Pruna, R., Ribas, J., Montoro, J. B., & Artells, R. (2015). The impact of single nucleotide polymorphisms on patterns of non-contact musculoskeletal soft tissue injuries in a football player population according to ethnicity. *Medicina Clinica*, *144*(3), 105–110. https://doi.org/10.1016/j.medcli.2013.09.026

Quinlan, K. G. R., Seto, J. T., Turner, N., Vandebrouck, A., Floetenmeyer, M., Macarthur, D. G., Raftery, J. M., Lek, M., Yang, N., Parton, R. G., Cooney, G. J., & North, K. N. (2010). Alpha-actinin-3 deficiency results in reduced glycogen phosphorylase activity and altered calcium handling in skeletal muscle. *Human Molecular Genetics*, *19*(7), 1335–1346. https://doi.org/10.1093/hmg/ddq010

Radnor, J. M., Staines, J., Bevan, J., Cumming, S. P., Kelly, A. L., Lloyd, R. S., & Oliver, J. L. (2021). Maturity has a greater association than relative age with physical performance in English male academy soccer players. *Sports*, *9*(12), Article 12. https://doi.org/10.3390/sports9120171

Rahim, M., Gibbon, A., Hobbs, H., van der Merwe, W., Posthumus, M., Collins, M., & September, A. V. (2014). The association of genes involved in the angiogenesis-associated signaling pathway with risk of anterior cruciate ligament rupture. *Journal of*

Orthopaedic Research: Official Publication of the Orthopaedic Research Society, *32*(12), 1612–1618. https://doi.org/10.1002/jor.22705

Rampinini, E., Impellizzeri, F. M., Castagna, C., Coutts, A. J., & Wisløff, U. (2009). Technical performance during soccer matches of the Italian Serie A league: Effect of fatigue and competitive level. *Journal of Science and Medicine in Sport*, *12*(1), 227–233. https://doi.org/10.1016/j.jsams.2007.10.002

Rankinen, T., & Bouchard, C. (2012). Gene-exercise interactions. *Progress in Molecular Biology and Translational Science*, *108*, 447–460. https://doi.org/10.1016/B978-0-12-398397-8.00017-4

Rees, T., Hardy, L., Güllich, A., Abernethy, B., Côté, J., Woodman, T., Montgomery, H., Laing, S., & Warr, C. (2016). The Great British Medalists Project: A review of current knowledge on the development of the world's best sporting talent. *Sports Medicine (Auckland, N.Z.)*, *46*(8), 1041–1058. https://doi.org/10.1007/s40279-016-0476-2

Reuter-Rice, K., Regier, M., Bennett, E., & Laskowitz, D. (2018). The effect of the relationship of APOE polymorphisms and cerebral vasospasm on functional outcomes in children with traumatic brain injury. *Biological Research for Nursing*, *20*(5), 566–576. https://doi.org/10.1177/1099800418785982

Rigat, B., Hubert, C., Alhenc-Gelas, F., Cambien, F., Corvol, P., & Soubrier, F. (1990). An insertion/deletion polymorphism in the angiotensin I-converting enzyme gene accounting for half the variance of serum enzyme levels. *The Journal of Clinical Investigation*, *86*(4), 1343–1346. https://doi.org/10.1172/JCI114844

Rizzo, M., Gensini, F., Fatini, C., Manetti, P., Pucci, N., Capalbo, A., Vono, M. C. R., & Galanti, G. (2003). ACE I/D polymorphism and cardiac adaptations in adolescent

athletes. *Medicine and Science in Sports and Exercise*, *35*(12), 1986–1990. https://doi.org/10.1249/01.MSS.0000098993.51693.0B

Roberts, A. H., Greenwood, D. A., Stanley, M., Humberstone, C., Iredale, F., & Raynor, A. (2019). Coach knowledge in talent identification: A systematic review and meta-synthesis. *Journal of Science and Medicine in Sport*, *22*(10), 1163–1172. https://doi.org/10.1016/j.jsams.2019.05.008

Rocha, A. W. de O., Nascimento, W. M. do, Oliveira, C. M. da C., Pereira Neto, J. M., Nascimento, O. V. do, Santos, J. O. L. dos, Viera, É. P., Brunetta, H. S., Pontes, I. da M., Astolfi Filho, S., Rocha, A. W. de O., Nascimento, W. M. do, Oliveira, C. M. da C., Pereira Neto, J. M., Nascimento, O. V. do, Santos, J. O. L. dos, Viera, É. P., Brunetta, H. S., Pontes, I. da M., & Astolfi Filho, S. (2020). Frequency of gene ACE I polymorphism I-D in athletes of different sports. *Revista Brasileira de Medicina Do Esporte*, *26*(2), 107–112. https://doi.org/10.1590/1517-869220202602218862

Rodas, G., Moreno-Pérez, V., Del Coso, J., Florit, D., Osaba, L., & Lucia, A. (2021). Alpha-Actinin-3 deficiency might affect recovery from non-contact muscle injuries: Preliminary findings in a top-level soccer team. *Genes*, *12*(5), 769. https://doi.org/10.3390/genes12050769

Rodas, G., Osaba, L., Arteta, D., Pruna, R., Fernández, D., & Lucia, A. (2019). Genomic prediction of tendinopathy risk in elite team sports. *International Journal of Sports Physiology and Performance*, *15*(4), 1–7. https://doi.org/10.1123/ijspp.2019-0431

Roth, S. M., Walsh, S., Liu, D., Metter, E. J., Ferrucci, L., & Hurley, B. F. (2008). The ACTN3 R577X nonsense allele is under-represented in elite-level strength athletes. *European*

Journal of Human Genetics : EJHG, *16*(3), 391–394. https://doi.org/10.1038/sj.ejhg.5201964

Ruiz, J. R., Arteta, D., Buxens, A., Artieda, M., Gómez-Gallego, F., Santiago, C., Yvert, T., Morán, M., & Lucia, A. (2010). Can we identify a power-oriented polygenic profile? *Journal of Applied Physiology (Bethesda, Md.: 1985)*, *108*(3), 561–566. https://doi.org/10.1152/japplphysiol.01242.2009

Ruiz, J. R., Buxens, A., Artieda, M., Arteta, D., Santiago, C., Rodríguez-Romo, G., Lao, J. I., Gómez-Gallego, F., & Lucia, A. (2010). The -174 G/C polymorphism of the IL6 gene is associated with elite power performance. *Journal of Science and Medicine in Sport*, *13*(5), 549–553. https://doi.org/10.1016/j.jsams.2009.09.005

Ruiz, J. R., Fernández del Valle, M., Verde, Z., Díez-Vega, I., Santiago, C., Yvert, T., Rodríguez-Romo, G., Gómez-Gallego, F., Molina, J. J., & Lucia, A. (2011). ACTN3 R577X polymorphism does not influence explosive leg muscle power in elite volleyball players. *Scandinavian Journal of Medicine & Science in Sports*, *21*(6), e34-41. https://doi.org/10.1111/j.1600-0838.2010.01134.x

Saber-Ayad, M. M., Nassar, Y. S., & Latif, I. A. (2014). Angiotensin-converting enzyme I/D gene polymorphism affects early cardiac response to professional training in young footballers. *Journal of the Renin-Angiotensin-Aldosterone System: JRAAS*, *15*(3), 236–242. https://doi.org/10.1177/1470320312471150

Sagoo, G. S., Little, J., & Higgins, J. P. T. (2009). Systematic reviews of genetic association studies. Human Genome Epidemiology Network. *PLoS Medicine*, *6*(3), e28. https://doi.org/10.1371/journal.pmed.1000028

Salanti, G., Amountza, G., Ntzani, E. E., & Ioannidis, J. P. A. (2005). Hardy-Weinberg equilibrium in genetic association studies: An empirical evaluation of reporting, deviations, and power. *European Journal of Human Genetics: EJHG*, *13*(7), 840–848. https://doi.org/10.1038/sj.ejhg.5201410

Salgueirosa, F. de M., Rodrigues, P., Seniski, G. G., Wharton, L., & Osiecki, R. (2017). ACTN3 R577X and ACE I/D genotype frequencies of professional soccer players in brazil. *Journal of Exercise Physiology Online*, *20*(6), 129–139.

Saloum de Neves Manta, F., Pereira, R., Vianna, R., Rodolfo Beuttenmüller de Araújo, A., Leite Góes Gitaí, D., Aparecida da Silva, D., de Vargas Wolfgramm, E., da Mota Pontes, I., Ivan Aguiar, J., Ozório Moraes, M., Fagundes de Carvalho, E., & Gusmão, L. (2013). Revisiting the genetic ancestry of Brazilians using autosomal AIM-Indels. *PloS One*, *8*(9), e75145. https://doi.org/10.1371/journal.pone.0075145

Sanchez-Roige, S., Gray, J. C., MacKillop, J., Chen, C.-H., & Palmer, A. A. (2018). The genetics of human personality. *Genes, Brain, and Behavior*, *17*(3), e12439. https://doi.org/10.1111/gbb.12439

Santiago, C., González-Freire, M., Serratosa, L., Morate, F. J., Meyer, T., Gómez-Gallego, F., & Lucia, A. (2008). ACTN3 genotype in professional soccer players. *British Journal of Sports Medicine*, *42*(1), 71–73. https://doi.org/10.1136/bjsm.2007.039172

Santoro, G. F., Mello, K. D. de, Oliveira Netto, Z. C. de, Pfutzenreuter, G., Bassan, J. C., Salgueirosa, F. de M., Santoro, G. F., Mello, K. D. de, Oliveira Netto, Z. C. de, Pfutzenreuter, G., Bassan, J. C., & Salgueirosa, F. de M. (2019). The influence of ACE I/D gene polymorphism in amateur American football athletes in Brazil. *Revista*

Brasileira de Medicina Do Esporte, *25*(6), 460–463. https://doi.org/10.1590/1517-869220192506198909

Sarmento, H., Anguera, M. T., Pereira, A., & Araújo, D. (2018). Talent identification and development in male football: A systematic review. *Sports Medicine*, *48*(4), 907–931. https://doi.org/10.1007/s40279-017-0851-7

Saward, C., Hulse, M., Morris, J. G., Goto, H., Sunderland, C., & Nevill, M. E. (2020). Longitudinal physical development of future professional male soccer players: Implications for talent identification and development? *Frontiers in Sports and Active Living*, *2*, 578203. https://doi.org/10.3389/fspor.2020.578203

Scanavini, D., Bernardi, F., Castoldi, E., Conconi, F., & Mazzoni, G. (2002). Increased frequency of the homozygous II ACE genotype in Italian Olympic endurance athletes. *European Journal of Human Genetics*, *10*(10), Article 10. https://doi.org/10.1038/sj.ejhg.5200852

Schneider, B. P., Radovich, M., Sledge, G. W., Robarge, J. D., Li, L., Storniolo, A. M., Lemler, S., Nguyen, A. T., Hancock, B. A., Stout, M., Skaar, T., & Flockhart, D. A. (2008). Association of polymorphisms of angiogenesis genes with breast cancer. *Breast Cancer Research and Treatment*, *111*(1), 157–163. https://doi.org/10.1007/s10549-007-9755-9

Schutte, N. M., Nederend, I., Hudziak, J. J., Bartels, M., & de Geus, E. J. C. (2016). Twin-sibling study and meta-analysis on the heritability of maximal oxygen consumption. *Physiological Genomics*, *48*(3), 210–219. https://doi.org/10.1152/physiolgenomics.00117.2015

Scott, M., & Kelso, P. (2008, April 25). *One club wants to use a gene-test to spot the new Ronaldo. Is this football's future?* The Guardian. http://www.theguardian.com/football/2008/apr/26/genetics

Scott, R. A., Irving, R., Irwin, L., Morrison, E., Charlton, V., Austin, K., Tladi, D., Deason, M., Headley, S. A., Kolkhorst, F. W., Yang, N., North, K., & Pitsiladis, Y. P. (2010). ACTN3 and ACE genotypes in elite Jamaican and US sprinters. *Medicine and Science in Sports and Exercise*, *42*(1), 107–112. https://doi.org/10.1249/MSS.0b013e3181ae2bc0

Serrano, A. L., Baeza-Raja, B., Perdiguero, E., Jardí, M., & Muñoz-Cánoves, P. (2008). Interleukin-6 is an essential regulator of satellite cell-mediated skeletal muscle hypertrophy. *Cell Metabolism*, *7*(1), 33–44. https://doi.org/10.1016/j.cmet.2007.11.011

Sessa, F., Chetta, M., Petito, A., Franzetti, M., Bafunno, V., Pisanelli, D., Sarno, M., Iuso, S., & Margaglione, M. (2011). Gene polymorphisms and sport attitude in Italian athletes. *Genetic Testing and Molecular Biomarkers*, *15*(4), 285–290. https://doi.org/10.1089/gtmb.2010.0179

Seto, J. T., Lek, M., Quinlan, K. G. R., Houweling, P. J., Zheng, X. F., Garton, F., MacArthur, D. G., Raftery, J. M., Garvey, S. M., Hauser, M. A., Yang, N., Head, S. I., & North, K. N. (2011). Deficiency of α-actinin-3 is associated with increased susceptibility to contraction-induced damage and skeletal muscle remodeling. *Human Molecular Genetics*, *20*(15), 2914–2927. https://doi.org/10.1093/hmg/ddr196

Seto, J. T., Quinlan, K. G. R., Lek, M., Zheng, X. F., Garton, F., MacArthur, D. G., Hogarth, M. W., Houweling, P. J., Gregorevic, P., Turner, N., Cooney, G. J., Yang, N., & North, K. N. (2013). ACTN3 genotype influences muscle performance through the regulation

of calcineurin signaling. *The Journal of Clinical Investigation*, *123*(10), 4255–4263. https://doi.org/10.1172/JCI67691

Shenoy, S., Tandon, S., Sandhu, J., & Bhanwer, A. S. (2010). Association of angiotensin converting enzyme gene polymorphism and Indian army triathletes performance. *Asian Journal of Sports Medicine*, *1*(3), 143–150. https://doi.org/10.5812/asjsm.34855

Sieghartsleitner, R., Zuber, C., Zibung, M., & Conzelmann, A. (2019). Science or coaches' eye? - Both! Beneficial collaboration of multidimensional measurements and coach assessments for efficient talent selection in elite youth football. *Journal of Sports Science & Medicine*, *18*(1), 32–43.

Silventoinen, K., Magnusson, P. K. E., Tynelius, P., Kaprio, J., & Rasmussen, F. (2008). Heritability of body size and muscle strength in young adulthood: A study of one million Swedish men. *Genetic Epidemiology*, *32*(4), 341–349. https://doi.org/10.1002/gepi.20308

Slimani, M., & Nikolaidis, P. T. (2019). Anthropometric and physiological characteristics of male soccer players according to their competitive level, playing position and age group: A systematic review. *The Journal of Sports Medicine and Physical Fitness*, *59*(1), 141–163. https://doi.org/10.23736/S0022-4707.17.07950-6

Smith, C., Graham, D. I., Murray, L. S., Stewart, J., & Nicoll, J. A. R. (2006). Association of APOE e4 and cerebrovascular pathology in traumatic brain injury. *Journal of Neurology, Neurosurgery, and Psychiatry*, *77*(3), 363–366. https://doi.org/10.1136/jnnp.2005.074617

Solé, X., Guinó, E., Valls, J., Iniesta, R., & Moreno, V. (2006). SNPStats: A web tool for the analysis of association studies. *Bioinformatics (Oxford, England)*, *22*(15), 1928–1929. https://doi.org/10.1093/bioinformatics/btl268

Staal, W. G. (2015). Autism, DRD3 and repetitive and stereotyped behavior, an overview of the current knowledge. *European Neuropsychopharmacology: The Journal of the European College of Neuropsychopharmacology*, *25*(9), 1421–1426. https://doi.org/10.1016/j.euroneuro.2014.08.011

Stang, A. (2010). Critical evaluation of the Newcastle-Ottawa scale for the assessment of the quality of nonrandomized studies in meta-analyses. *European Journal of Epidemiology*, *25*(9), 603–605. https://doi.org/10.1007/s10654-010-9491-z

Stastny, P., Lehnert, M., De Ste Croix, M., Petr, M., Svoboda, Z., Maixnerova, E., Varekova, R., Botek, M., Petrek, M., Kocourkova, L., & Cięszczyk, P. (2019). Effect of COL5A1, GDF5, and PPARA genes on a movement screen and neuromuscular performance in adolescent team sport athletes. *Journal of Strength and Conditioning Research*, *33*(8), 2057–2065. https://doi.org/10.1519/JSC.0000000000003142

Steca, P., Baretta, D., Greco, A., D'Addario, M., & Monzani, D. (2018). Associations between personality, sports participation and athletic success. A comparison of Big Five in sporting and non-sporting adults. *Personality and Individual Differences*, *121*, 176–183. https://doi.org/10.1016/j.paid.2017.09.040

Stølen, T., Chamari, K., Castagna, C., & Wisløff, U. (2005). Physiology of soccer: An update. *Sports Medicine (Auckland, N.Z.)*, *35*(6), 501–536. https://doi.org/10.2165/00007256-200535060-00004

Strawbridge, R. J., Ward, J., Cullen, B., Tunbridge, E. M., Hartz, S., Bierut, L., Horton, A., Bailey, M. E. S., Graham, N., Ferguson, A., Lyall, D. M., Mackay, D., Pidgeon, L. M., Cavanagh, J., Pell, J. P., O'Donovan, M., Escott-Price, V., Harrison, P. J., & Smith, D. J. (2018). Genome-wide analysis of self-reported risk-taking behaviour and cross-disorder genetic correlations in the UK Biobank cohort. *Translational Psychiatry*, *8*(1), 39. https://doi.org/10.1038/s41398-017-0079-1

Strittmatter, W. J., Weisgraber, K. H., Huang, D. Y., Dong, L. M., Salvesen, G. S., Pericak-Vance, M., Schmechel, D., Saunders, A. M., Goldgaber, D., & Roses, A. D. (1993). Binding of human apolipoprotein E to synthetic amyloid beta peptide: Isoform-specific effects and implications for late-onset Alzheimer disease. *Proceedings of the National Academy of Sciences of the United States of America*, *90*(17), 8098–8102.

Swann, C., Moran, A., & Piggott, D. (2015). Defining elite athletes: Issues in the study of expert performance in sport psychology. *Psychology of Sport and Exercise*, *16*, 3–14. https://doi.org/10.1016/j.psychsport.2014.07.004

Tanaka, M., Wang, G., & Pitsiladis, Y. P. (2016). Advancing sports and exercise genomics: Moving from hypothesis-driven single study approaches to large multi-omics collaborative science. *Physiological Genomics*, *48*(3), 173–174. https://doi.org/10.1152/physiolgenomics.00009.2016

Tanisawa, K., Wang, G., Seto, J., Verdouka, I., Twycross-Lewis, R., Karanikolou, A., Tanaka, M., Borjesson, M., Luigi, L. D., Dohi, M., Wolfarth, B., Swart, J., Bilzon, J. L. J., Badtieva, V., Papadopoulou, T., Casasco, M., Geistlinger, M., Bachl, N., Pigozzi, F., & Pitsiladis, Y. (2020). Sport and exercise genomics: The FIMS 2019 consensus statement

update. *British Journal of Sports Medicine*, *54*(16), 969–975. https://doi.org/10.1136/bjsports-2019-101532

Tashakkori, A., & Teddlie, C. (1998). *Mixed methodology: Combining qualitative and quantitative approaches* (pp. xi, 185). Sage Publications.

Teddlie, C., & Tashakkori, A. (Eds.). (2009). *Foundations of mixed methods research*. Sage Publications.

Ten Hagen, K. G., Fritz, T. A., & Tabak, L. A. (2003). All in the family: The UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferases. *Glycobiology*, *13*(1), 1R-16R. https://doi.org/10.1093/glycob/cwg007

Terrell, T. R., Abramson, R., Barth, J. T., Bennett, E., Cantu, R. C., Sloane, R., Laskowitz, D. T., Erlanger, D. M., McKeag, D., Nichols, G., Valentine, V., & Galloway, L. (2018). Genetic polymorphisms associated with the risk of concussion in 1056 college athletes: A multicentre prospective cohort study. *British Journal of Sports Medicine*, *52*(3), 192–198. https://doi.org/10.1136/bjsports-2016-097419

Terrell, T. R., Bostick, R. M., Abramson, R., Xie, D., Barfield, W., Cantu, R., Stanek, M., & Ewing, T. (2008). APOE, APOE promoter, and Tau genotypes and risk for concussion in college athletes. *Clinical Journal of Sport Medicine: Official Journal of the Canadian Academy of Sport Medicine*, *18*(1), 10–17. https://doi.org/10.1097/JSM.0b013e31815c1d4c

Tharabenjasin, P., Pabalan, N., & Jarjanazi, H. (2019). Association of the ACTN3 R577X (rs1815739) polymorphism with elite power sports: A meta-analysis. *PloS One*, *14*(5), e0217390. https://doi.org/10.1371/journal.pone.0217390

Thompson, W. R., & Binder-Macleod, S. A. (2006). Association of genetic factors With selected measures of physical performance. *Physical Therapy*, *86*(4), 585–591.

Tierney, R. T., Mansell, J. L., Higgins, M., McDevitt, J. K., Toone, N., Gaughan, J. P., Mishra, A., & Krynetskiy, E. (2010). Apolipoprotein E genotype and concussion in college athletes. *Clinical Journal of Sport Medicine: Official Journal of the Canadian Academy of Sport Medicine*, *20*(6), 464–468. https://doi.org/10.1097/JSM.0b013e3181fc0a81

Toma, C., Hervás, A., Balmaña, N., Salgado, M., Maristany, M., Vilella, E., Aguilera, F., Orejuela, C., Cuscó, I., Gallastegui, F., Pérez-Jurado, L. A., Caballero-Andaluz, R., Diego-Otero, Y. de, Guzmán-Alvarez, G., Ramos-Quiroga, J. A., Ribasés, M., Bayés, M., & Cormand, B. (2013). Neurotransmitter systems and neurotrophic factors in autism: Association study of 37 genes suggests involvement of DDC. *The World Journal of Biological Psychiatry: The Official Journal of the World Federation of Societies of Biological Psychiatry*, *14*(7), 516–527. https://doi.org/10.3109/15622975.2011.602719

Tucker, R., & Collins, M. (2012). What makes champions? A review of the relative contribution of genes and training to sporting success. *British Journal of Sports Medicine*, *46*(8), 555–561. https://doi.org/10.1136/bjsports-2011-090548

Tunbridge, E. M., Narajos, M., Harrison, C. H., Beresford, C., Cipriani, A., & Harrison, P. J. (2019). Which dopamine polymorphisms are functional? Systematic review and meta-analysis of COMT, DAT, DBH, DDC, DRD1-5, MAOA, MAOB, TH, VMAT1, and VMAT2. *Biological Psychiatry*, *86*(8), 608–620. https://doi.org/10.1016/j.biopsych.2019.05.014

Turgut, G., Turgut, S., Genc, O., Atalay, A., & Atalay, E. O. (2004). The angiotensin converting enzyme I/D polymorphism in Turkish athletes and sedentary controls. *Acta Medica (Hradec Kralove)*, *47*(2), 133–136.

Turkheimer, E. (2000). Three laws of behavior genetics and what they mean. *Current Directions in Psychological Science*, *9*(5), 160–164. https://doi.org/10.1111/1467-8721.00084

Ullén, F., Hambrick, D. Z., & Mosing, M. A. (2016). Rethinking expertise: A multifactorial gene-environment interaction model of expert performance. *Psychological Bulletin*, *142*(4), 427–446. https://doi.org/10.1037/bul0000033

Ulucan, K., Bayyurt, G. M., Konuk, M., & Güney, A. I. (2014). Effect of alpha-actinin-3 gene on trained and untrained Turkish middle-school children's sprinting performance: A pilot study. *Biological Rhythm Research*, *45*(4), 509–514. https://doi.org/10.1080/09291016.2013.867628

Ulucan, K., Sercan, C., & Biyikli, T. (2015). Distribution of angiotensin-1 converting enzyme insertion/deletion and α-actinin-3 codon 577 polymorphisms in Turkish male soccer players. *Genetics & Epigenetics*, *7*, 1–4. https://doi.org/10.4137/GEG.S31479

Vaeyens, R., Lenoir, M., Williams, A. M., & Philippaerts, R. M. (2008). Talent identification and development programmes in sport: Current models and future directions. *Sports Medicine*, *38*(9), 703–714. https://doi.org/10.2165/00007256-200838090-00001

Vaeyens, R., Malina, R. M., Janssens, M., Van Renterghem, B., Bourgois, J., Vrijens, J., & Philippaerts, R. M. (2006). A multidisciplinary selection model for youth soccer: The Ghent Youth Soccer Project. *British Journal of Sports Medicine*, *40*(11), 928–934; discussion 934. https://doi.org/10.1136/bjsm.2006.029652

Valeeva, E. V., Ahmetov, I. I., & Rees, T. (2019). Chapter Seven—Psychogenetics and sport. In D. Barh & I. I. Ahmetov (Eds.), *Sports, Exercise, and Nutritional Genomics* (pp. 147–165). Academic Press. https://doi.org/10.1016/B978-0-12-816193-7.00007-5

Varillas Delgado, D., Tellería Orriols, J. J., Monge Martín, D., & Del Coso, J. (2020). Genotype scores in energy and iron-metabolising genes are higher in elite endurance athletes than in nonathlete controls. *Applied Physiology, Nutrition, and Metabolism*, *45*(11), 1225–1231. https://doi.org/10.1139/apnm-2020-0174

Varley, I., Greeves, J. P., Sale, C., Friedman, E., Moran, D. S., Yanovich, R., Wilson, P. J., Gartland, A., Hughes, D. C., Stellingwerff, T., Ranson, C., Fraser, W. D., & Gallagher, J. A. (2016). Functional polymorphisms in the P2X7 receptor gene are associated with stress fracture injury. *Purinergic Signalling*, *12*(1), 103–113. https://doi.org/10.1007/s11302-016-9495-6

Varley, I., Hughes, D. C., Greeves, J. P., Fraser, W. D., & Sale, C. (2018). SNPs in the vicinity of P2X7R, RANK/RANKL/OPG and Wnt signalling pathways and their association with bone phenotypes in academy footballers. *Bone*, *108*, 179–185. https://doi.org/10.1016/j.bone.2018.01.007

Varley, I., Hughes, D. C., Greeves, J. P., Stellingwerff, T., Ranson, C., Fraser, W. D., & Sale, C. (2015). RANK/RANKL/OPG pathway: Genetic associations with stress fracture period prevalence in elite athletes. *Bone*, *71*, 131–136. https://doi.org/10.1016/j.bone.2014.10.004

Varley, I., Hughes, D. C., Greeves, J. P., Stellingwerff, T., Ranson, C., Fraser, W. D., & Sale, C. (2018). The association of novel polymorphisms with stress fracture injury in elite

athletes: Further insights from the SFEA cohort. *Journal of Science and Medicine in Sport*, *21*(6), 564–568. https://doi.org/10.1016/j.jsams.2017.10.038

Varley, I., Patel, S., Williams, A. G., & Hennis, P. J. (2018). The current use, and opinions of elite athletes and support staff in relation to genetic testing in elite sport within the UK. *Biology of Sport*, *35*(1), 13–19. https://doi.org/10.5114/biolsport.2018.70747

Varley, M. C., Gregson, W., McMillan, K., Bonanno, D., Stafford, K., Modonutti, M., & Salvo, V. D. (2017). Physical and technical performance of elite youth soccer players during international tournaments: Influence of playing position and team success and opponent quality. *Science and Medicine in Football*, *1*(1), 18–29. https://doi.org/10.1080/02640414.2016.1230676

Vaughn, N. H., Stepanyan, H., Gallo, R. A., & Dhawan, A. (2017). Genetic factors in tendon injury: A systematic review of the literature. *Orthopaedic Journal of Sports Medicine*, *5*(8). https://doi.org/10.1177/2325967117724416

Viechtbauer, W., & Cheung, M. W.-L. (2010). Outlier and influence diagnostics for meta-analysis. *Research Synthesis Methods*, *1*(2), 112–125. https://doi.org/10.1002/jrsm.11

Vieira, L. H. P., Carling, C., Barbieri, F. A., Aquino, R., & Santiago, P. R. P. (2019). Match running performance in young soccer players: A systematic review. *Sports Medicine (Auckland, N.Z.)*, *49*(2), 289–318. https://doi.org/10.1007/s40279-018-01048-8

Visscher, P. M., Brown, M. A., McCarthy, M. I., & Yang, J. (2012). Five years of GWAS discovery. *The American Journal of Human Genetics*, *90*(1), 7–24. https://doi.org/10.1016/j.ajhg.2011.11.029

Visscher, P. M., Medland, S. E., Ferreira, M. A. R., Morley, K. I., Zhu, G., Cornes, B. K., Montgomery, G. W., & Martin, N. G. (2006). Assumption-free estimation of heritability

from genome-wide identity-by-descent sharing between full siblings. *PLoS Genetics*, *2*(3), e41. https://doi.org/10.1371/journal.pgen.0020041

Vlahovich, N., Fricker, P. A., Brown, M. A., & Hughes, D. (2017). Ethics of genetic testing and research in sport: A position statement from the Australian Institute of Sport. *British Journal of Sports Medicine*, *51*(1), 5–11. https://doi.org/10.1136/bjsports-2016-096661

Von Elm, E., Altman, D. G., Egger, M., Pocock, S. J., Gøtzsche, P. C., Vandenbroucke, J. P., & STROBE Initiative. (2008). The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: Guidelines for reporting observational studies. *Journal of Clinical Epidemiology*, *61*(4), 344–349. https://doi.org/10.1016/j.jclinepi.2007.11.008

Waldén, M., Hägglund, M., Werner, J., & Ekstrand, J. (2011). The epidemiology of anterior cruciate ligament injury in football (soccer): A review of the literature from a gender-related perspective. *Knee Surgery, Sports Traumatology, Arthroscopy: Official Journal of the ESSKA*, *19*(1), 3–10. https://doi.org/10.1007/s00167-010-1172-7

Waldron, M., & Murphy, A. (2013). A comparison of physical abilities and match performance characteristics among elite and subelite under-14 soccer players. *Pediatric Exercise Science*, *25*(3), 423–434. https://doi.org/10.1123/pes.25.3.423

Wallace, J. L., & Norton, K. I. (2014). Evolution of World Cup soccer final games 1966-2010: Game structure, speed and play patterns. *Journal of Science and Medicine in Sport*, *17*(2), 223–228. https://doi.org/10.1016/j.jsams.2013.03.016

Wang, C., Li, H., Chen, K., Wu, B., & Liu, H. (2017). Association of polymorphisms rs1800012 in COL1A1 with sports-related tendon and ligament injuries: A meta-analysis. *Oncotarget*, *8*(16), 27627–27634. https://doi.org/10.18632/oncotarget.15271

Wang, G., Padmanabhan, S., Miyamoto-Mikami, E., Fuku, N., Tanaka, M., Miyachi, M., Murakami, H., Cheng, Y.-C., Mitchell, B. D., Austin, K. G., & Pitsiladis, Y. P. (2014). GWAS of elite Jamaican, African American and Japanese sprint athletes. *Medicine & Science in Sports & Exercise*, *46*(5), 596–598. https://doi.org/10.1249/01.mss.0000495260.97185.ec

Wang, G., Padmanabhan, S., Wolfarth, B., Fuku, N., Lucia, A., Ahmetov, I. I., Cieszczyk, P., Collins, M., Eynon, N., Klissouras, V., Williams, A., & Pitsiladis, Y. (2013). Genomics of elite sporting performance: What little we know and necessary advances. *Advances in Genetics*, *84*, 123–149. https://doi.org/10.1016/B978-0-12-407703-4.00004-9

Wang, H., & Keiser, J. A. (1998). Vascular endothelial growth factor upregulates the expression of matrix metalloproteinases in vascular smooth muscle cells: Role of flt-1. *Circulation Research*, *83*(8), 832–840. https://doi.org/10.1161/01.res.83.8.832

Webborn, N., Williams, A., McNamee, M., Bouchard, C., Pitsiladis, Y., Ahmetov, I., Ashley, E., Byrne, N., Camporesi, S., Collins, M., Dijkstra, P., Eynon, N., Fuku, N., Garton, F. C., Hoppe, N., Holm, S., Kaye, J., Klissouras, V., Lucia, A., … Wang, G. (2015). Direct-to-consumer genetic testing for predicting sports performance and talent identification: Consensus statement. *British Journal of Sports Medicine*, *49*(23), 1486–1491. https://doi.org/10.1136/bjsports-2015-095343

Wessner, B., Stuparits, P., Fail, C., Pavic, F., Tschan, H., & Bachl, N. (2016). Genetic polymorphisms in alpha-actinin 3 and adrenoceptor beta genes in Austrian elite athletes and healthy controls. *Schweizerische Zeitschrift Fur Sportmedizin Und Sporttraumatologie*, *64*, 13–19.

Weyerstraß, J., Stewart, K., Wesselius, A., & Zeegers, M. (2018). Nine genetic polymorphisms associated with power athlete status—A Meta-Analysis. *Journal of Science and Medicine in Sport*, *21*(2), 213–220. https://doi.org/10.1016/j.jsams.2017.06.012

Whitehead, A., & Whitehead, J. (1991). A general parametric approach to the meta-analysis of randomized clinical trials. *Statistics in Medicine*, *10*(11), 1665–1677. https://doi.org/10.1002/sim.4780101105

Williams, A. G., & Folland, J. P. (2008). Similarity of polygenic profiles limits the potential for elite human physical performance. *The Journal of Physiology*, *586*(1), 113–121. https://doi.org/10.1113/jphysiol.2007.141887

Williams, A. M., Ford, P. R., & Drust, B. (2020). Talent identification and development in soccer since the millennium. *Journal of Sports Sciences*, *38*(11–12), 1199–1210. https://doi.org/10.1080/02640414.2020.1766647

Williams, A. M., & Reilly, T. (2000). Talent identification and development in soccer. *Journal of Sports Sciences*, *18*(9), 657–667. https://doi.org/10.1080/02640410050120041

Yamin, C., Duarte, J. A. R., Oliveira, J. M. F., Amir, O., Sagiv, M., Eynon, N., Sagiv, M., & Amir, R. E. (2008). IL6 (-174) and TNFA (-308) promoter polymorphisms are associated with systemic creatine kinase response to eccentric exercise. *European Journal of Applied Physiology*, *104*(3), 579–586. https://doi.org/10.1007/s00421-008-0728-4

Yan, X., Papadimitriou, I., Lidor, R., & Eynon, N. (2016). Nature versus nurture in determining athletic ability. *Medicine and Sport Science*, *61*, 15–28. https://doi.org/10.1159/000445238

Yang, N., MacArthur, D. G., Gulbin, J. P., Hahn, A. G., Beggs, A. H., Easteal, S., & North, K. (2003). ACTN3 genotype is associated with human elite athletic performance. *American Journal of Human Genetics*, *73*(3), 627–631. https://doi.org/10.1086/377590

Yang, N., MacArthur, D. G., Wolde, B., Onywera, V. O., Boit, M. K., Lau, S. Y. M.-A., Wilson, R. H., Scott, R. A., Pitsiladis, Y. P., & North, K. (2007). The ACTN3 R577X polymorphism in East and West African athletes. *Medicine and Science in Sports and Exercise*, *39*(11), 1985–1988. https://doi.org/10.1249/mss.0b013e31814844c9

Yengo, L., Sidorenko, J., Kemper, K. E., Zheng, Z., Wood, A. R., Weedon, M. N., Frayling, T. M., Hirschhorn, J., Yang, J., Visscher, P. M., & GIANT Consortium. (2018). Meta-analysis of genome-wide association studies for height and body mass index in ~700000 individuals of European ancestry. *Human Molecular Genetics*, *27*(20), 3641–3649. https://doi.org/10.1093/hmg/ddy271

Young, K., Samiric, T., Feller, J., & Cook, J. (2011). Extracellular matrix content of ruptured anterior cruciate ligament tissue. *The Knee*, *18*(4), 242–246. https://doi.org/10.1016/j.knee.2010.05.008

Yvonne Feilzer, M. (2010). Doing mixed methods research pragmatically: Implications for the rediscovery of pragmatism as a research paradigm. *Journal of Mixed Methods Research*, *4*(1), 6–16. https://doi.org/10.1177/1558689809349691

Zabetian, C. P., Anderson, G. M., Buxbaum, S. G., Elston, R. C., Ichinose, H., Nagatsu, T., Kim, K. S., Kim, C. H., Malison, R. T., Gelernter, J., & Cubells, J. F. (2001). A quantitative-trait analysis of human plasma-dopamine beta-hydroxylase activity: Evidence for a major functional polymorphism at the DBH locus. *American Journal of Human Genetics*, *68*(2), 515–522. https://doi.org/10.1086/318198

Zempo, H., Miyamoto-Mikami, E., Kikuchi, N., Fuku, N., Miyachi, M., & Murakami, H. (2017). Heritability estimates of muscle strength-related phenotypes: A systematic review and meta-analysis. *Scandinavian Journal of Medicine & Science in Sports*, *27*(12), 1537–1546. https://doi.org/10.1111/sms.12804

Zhang, B., Tanaka, H., Shono, N., Miura, S., Kiyonaga, A., Shindo, M., & Saku, K. (2003). The I allele of the angiotensin-converting enzyme gene is associated with an increased percentage of slow-twitch type I fibers in human skeletal muscle. *Clinical Genetics*, *63*(2), 139–144. https://doi.org/10.1034/j.1399-0004.2003.00029.x

Zhao, M., Hu, Y., Yu, Y., Lin, Q., Yang, J., Su, S. B., Xu, G.-T., & Yang, T. (2016). Involvement of IL-37 in the pathogenesis of proliferative diabetic retinopathy. *Investigative Ophthalmology & Visual Science*, *57*(7), 2955–2962. https://doi.org/10.1167/iovs.15-18505

Zhou, C., Gómez, M.-Á., & Lorenzo, A. (2020). The evolution of physical and technical performance parameters in the Chinese Soccer Super League. *Biology of Sport*, *37*(2), 139–145. https://doi.org/10.5114/biolsport.2020.93039

Zhou, W., Xu, D., Peng, X., Zhang, Q., Jia, J., & Crutcher, K. A. (2008). Meta-analysis of APOE4 allele and outcome after traumatic brain injury. *Journal of Neurotrauma*, *25*(4), 279–290. https://doi.org/10.1089/neu.2007.0489

Zychowska, M., Półrola, P., Chruściński, G., Zielińska, J., & Góral-Półrola, J. (2017). Effects of sauna bathing on stress-related genes expression in athletes and non-athletes. *Annals of Agricultural and Environmental Medicine: AAEM*, *24*(1), 104–107. https://doi.org/10.5604/12321966.1233977

12 Appendices

12.1 Appendix A: Survey template

1. Which of the following best describes your primary job role?

- ☐ Athlete
- ☐ Coach
- ☐ Doctor
- ☐ Nutritionist
- ☐ Performance Analyst
- ☐ Physiotherapist
- ☐ Sport Scientist
- ☐ Strength and Conditioning Coach

If other, please specify: _____

2. Which age range do you belong to?

- ☐ 18-24
- ☐ 25-34
- ☐ 35-44
- ☐ 45-54
- ☐ 55-64
- ☐ 65+

3. Which sex are you?

- ☐ Male
- ☐ Female
- ☐ Prefer not to say

If other, please specify: _____

4. What level of football do you currently work/play at?

☐ Senior

☐ Academy

5. Which division is your club currently playing in?

☐ 1st Division (i.e., Premier League, La Liga, Serie A)
☐ 2nd Division (i.e., EFL Championship, Segunda Division, Serie B)
☐ 3rd Division (i.e., EFL League 1, Segunda Division B, Serie C)
☐ 4th Division (i.e., EFL League 2)
☐ Non-league (i.e., National League, Tercera Division, Serie D)

If other, please specify: _____

6. In your opinion, to what extent are **genetics responsible** for the following:

	Largely	Moderately	Slightly	No influence	Unsure
Strength	☐	☐	☐	☐	☐
Power	☐	☐	☐	☐	☐
Endurance	☐	☐	☐	☐	☐
Technical ability	☐	☐	☐	☐	☐
Decision making	☐	☐	☐	☐	☐
Resilience	☐	☐	☐	☐	☐
Personality	☐	☐	☐	☐	☐
Injury	☐	☐	☐	☐	☐

Please feel free to comment further: _____

7. Are you or have you ever been at a **club or organisation** that have used genetic testing to:

   a) aid performance

   ☐ Yes

   ☐ No

   b) mitigate injury risk

   ☐ Yes

   ☐ No

Please feel free to comment further:

_____

8. Have you ever utilised genetic testing **in the past** to:

   a) aid performance

   ☐ Yes

   ☐ No

   b) mitigate injury risk

   ☐ Yes

   ☐ No

Please feel free to comment further:

_____

9. Would you consider utilising genetic tests **in the future** to:

b) aid performance

☐ Yes

☐ No

b) mitigate injury risk

☐ Yes

☐ No

Please feel free to comment further:

_____

10. Genetic testing is **beneficial** and has great utility in sport:

	Strongly agree	Agree	Unsure	Disagree	Strongly Disagree
Currently	☐	☐	☐	☐	☐
In the future	☐	☐	☐	☐	☐

Please feel free to comment further:

_____

11. Genetic testing in sport **<u>should be used</u>** for the following:

	Strongly agree	Agree	Unsure	Disagree	Strongly Disagree
Talent identification/selection	☐	☐	☐	☐	☐
Athlete development/training	☐	☐	☐	☐	☐
Injury risk/prevention	☐	☐	☐	☐	☐

Please feel free to comment further:

_____

12. In your opinion, to what extent do the following factors **limit** the implementation of genetic testing in sport?

	Largely	Moderately	Slightly	No influence	Unsure
Cost	☐	☐	☐	☐	☐
Time	☐	☐	☐	☐	☐
Knowledge/inability to interpret results	☐	☐	☐	☐	☐
Fear of data privacy issues	☐	☐	☐	☐	☐
Culture/tradition	☐	☐	☐	☐	☐
Ethical issues	☐	☐	☐	☐	☐
Media reaction	☐	☐	☐	☐	☐

Please feel free to comment further:

_____

13. How much would you **expect to pay** for the following:

	£0	£1-£25	£26-£50	£51-£75	£76-£100	£101-£150	£151-£200	£201-£250
Genetic test	☐	☐	☐	☐	☐	☐	☐	☐
Consultancy to analyse & interpret results (per hour)	☐	☐	☐	☐	☐	☐	☐	☐
Genetic counselling session to explain results (per hour)	☐	☐	☐	☐	☐	☐	☐	☐
Education/training for staff, players, parents (per hour)	☐	☐	☐	☐	☐	☐	☐	☐

Please feel free to comment further:

_____

14. In your opinion, who should be allowed to **access and/or use genetic information** on athletes? (Please select all that apply)

- ☐ Athlete
- ☐ Genetic testing company
- ☐ National government
- ☐ Parents
- ☐ Scientific staff
- ☐ Sports club
- ☐ Sports league/association
- ☐ Anyone athlete consents for data sharing

Please feel free to comment further:

_____

15. To what extent have the following sources of information **influenced your opinions** regarding genetic associations and genetic testing in sport?

	Mostly	Largely	Moderately	Slightly	No influence
Scientific journal articles/conferences	☐	☐	☐	☐	☐
Professional sport organisations	☐	☐	☐	☐	☐
Television or press advertisement	☐	☐	☐	☐	☐
Online videos	☐	☐	☐	☐	☐
Social media	☐	☐	☐	☐	☐
Word of mouth	☐	☐	☐	☐	☐
Celebrity endorsement	☐	☐	☐	☐	☐

Please feel free to comment further:

_____

16. In your opinion, do you think **you have sufficient knowledge** of the relationship between genetics and performance/injury risk to make an informed decision on whether genetic testing should take place within your club?

☐ Yes
☐ No

Please feel free to comment further:

_____

17. In your opinion, do you think **other key stakeholders** at your club/organisation have sufficient knowledge of the relationship between genetics and performance/injury risk to make an informed decision on whether genetic testing should take place within your club?

☐ Yes
☐ No

Please feel free to comment further:

_____

18. In your opinion, is genetic research **communicated effectively** to athletes and coaches for them to make informed decisions regarding genetic testing?

☐ Yes
☐ No

Please feel free to comment further:

_____

19. In your opinion, are **educational workshops** required at your club regarding genetic research in sport?

☐ Yes
☐ No

Please feel free to comment further:

_____

20. In your opinion, is the **regular opportunity** to speak to a genetic specialist or the addition of a genetic consultant at your club required?

☐ Yes
☐ No

Please feel free to comment further:

_____

21. Would you personally be **interested in learning more** about the contribution of genetics to sporting performance/injury risk and the validity of genetic testing?

☐ Yes
☐ No

Please feel free to comment further:

_____