AN INVESTIGATION OF GENETIC FACTORS IN SOFT TISSUE INJURIES AND MUSCLE TYPES WITHIN FOOTBALL PLAYERS

A Thesis Submitted to
the Graduate School of Engineering and Sciences of
Izmir Institute of Technology
in Partial Fulfillment of the Requirements for the Degree of

MASTER OF SCIENCE

in Biotechnology

by
Yasemin ÜK

May 2022
İZMİR

ACKNOWLEDGEMENTS

I would like to thank my thesis advisor, Assoc. Prof. Efe Sezgin for his constant support, patience, guidance and encouraging approach through the learning and writing process of my master thesis.

Secondly, I would like to thank all the members of Uniqgene team; Dudu Seher Yurt, Irmak Ege and Oğuzhan Doğan for support and assistance throughout the journey of experiencing new technologies, methods and R&D progresses that helped me to learn more about sports genetics. Thanks to their making work entertaining, I didn't feel like I was working at all. Also, I would like to thank Berkay Bozkurt for his assistance in laboratory designs, experiments, and troubleshooting guidance through the wet-lab processes.

Most importantly, I wish to thank my family who deserves the better expression of my gratefulness for their endless support and prayers. This thesis is dedicated to my dear mother, Hacer Esin Yalınkılınç who supported me and affectionately raised me for the whole of my life. Also, due to her illness in breast cancer when I was at middle school encouraged me to investigate the mechanism of diseases and came out as an important part of a journey that made me interested in molecular biology and deriven me to search for solutions. I would like to thank my aunts, Tülin Yalınkılıç and Zühre Yalınkılıç, is the ones who always helped me, treated me as their children and supported me throughout my entire college life.

Last but not least, very special thanks go also to the many others that go unmentioned but have contributed in one way or another to the successful outcome of this work.

ABSTRACT

AN INVESTIGATION OF GENETIC FACTORS IN SOFT TISSUE INJURIES AND MUSCLE TYPES WITHIN FOOTBALL PLAYERS

When it comes to athletic performance, genes that affect many factors such as increased aerobic capacity, gene regions associated with strength and athletic ability, oxygen capacity, muscle structure, energy metabolism, and gene variants that cause these genes to be affected at different levels in each individual have been the subject of research for a long time. is addressed.

Amateur and professional athletes train according to their injury history after the problems they encounter. However, such measures are insufficient to prevent new injuries in the long run and ensure that the athlete has a healthy season. Similarly, anthropometric tests used in athlete selection and systems that are insufficient to process the data are insufficient to predict the physiological characteristics of this individual in the future and the development of the athlete.

In this research, it is aimed to associate detailed genetic parameters by including the genetic marker of the athlete that affects a successful and healthy sports life, and to inform the trainers by reporting the genes that are statistically meaningful with physical data; In this direction, it is aimed to develop and widely provide new test technologies. The aim is to test the hypothesis that athletes who have been genetically scored and matched with the right training accordingly show more improvement. With this setup, the project will be one of the comprehensive studies in sports genetics studies conducted in Turkey, which includes scoring, enables quantitative measurements, and combines genetic factors with physical tests in the international literature.

Keywords: *Sports Genetics, Soft Tissue Injury, Athletic Performance*

ÖZET

FUTBOL OYUNCULARINDAKİ YUMUŞAK DOKU YARALANMALARINDA VE KAS TİPLERİNDE GENETİK FAKTÖRLERİN İNCELENMESİ

Atletik performans söz konusu edildiğinde, arttırılmış aerobik kapasite, güç ve atletik kabiliyetle ilişkilendirilen gen bölgeleri, oksijen kapasitesi, kas yapısı, enerji metabolizması gibi pek çok faktörü etkileyen genler ve bu genlerin her bireyde farklı düzeyde etkilenmesine neden olan gen varyantları, uzun zamandır araştırma konusu olarak ele alınmaktadır. Amatör ve profesyonel sporcular, karşılaştıkları sorunların ardından sakatlık geçmişlerine göre antrenman yapmaktadır. Ancak, bu tür önlemler uzun vadede yeni sakatlıkları önlemekte ve sporcunun sağlıklı olarak bir sezon geçirmesini sağlamakta yetersiz kalmaktadır. Benzer şekilde, sporcu seçimlerinde kullanılan antropometrik testler, verileri işlemekte yetersiz kalan sistemler, ileride bu bireyin nasıl fizyolojik özellikleri olacağını ve sporcunun gelişimini öngörmekte yetersiz kalmaktadır.

Bu araştırmada, sporcunun başarılı ve sağlıklı bir spor yaşantısına etki eden genetik markırı dahil edilerek detaylı genetik parametrelerin ilişkilendirilmesi ve fiziksel veriler ile istatistiksel olarak anlamlandırılan genlerin raporlanarak antrenörlerin bilgilendirilmesi; bu doğrultuda yeni test teknolojilerinin geliştirilmesi ve yaygın olarak sağlanması amaçlanmaktadır. Genetik olarak skorlaması yapılmış ve bu doğrultuda doğru antrenman ile eşleştirilmiş sporcuların daha fazla gelişme gösterdiği hipotezini test etmektir. Bu kurgusuyla proje, Türkiye'de yapılan spor genetiği çalışmalarında, skorlamaları dahil eden, kantitatif ölçümlere olanak tanıyan ve uluslararası literatürde de genetik faktörleri fiziksel testler ile birleştiren kapsamlı çalışmalardan bir tanesi olacaktır.

Anahtar Kelimeler: *Spor Genetiği, Yumuşak Doku Yaralanmaları, Atletik Performans*

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 Genetics

1.1.1 Sports genetics

Genetics is a science that continues to develop rapidly within the fields of health and has been affecting our world with its wide spectrum since its foundation. A gene is a unit of heredity, defined as the nucleotide sequence that makes up a particular part of a chromosome. In other words, gene is also used in the sense of 'the biological unit that carries a certain characteristic passed from a parent to its children' (Pearson 2006). A wide variety of factors determine athletes' success, including genetics, epigenetics, training, nutrition, motivation, and environmental conditions. Since physical performance is a complex that includes the effects of environmental factors (nutrition, lifestyle, climate etc.) and genetics, every aspect should be evaluated for the best performance (Ahmetov et al. 2015).

To date, the effect of genetics on sports performance has been recorded to be between 30% and 80%, out of 120 variants affecting sports performance, 77 were found to be associated with susceptibility to endurance sports, while studies conducted with elite athletes of different populations revealed that 43 markers were associated with power sports (Ahmetov et al. 2016). These variants were revealed by examining the variants in gene regions that affect many features such as muscle type, muscle strength and mass, ion balance in the body, lipid and energy metabolism, vascular formation, and blood circulation, VO2max, collagen production, and muscle inflammation (Pickering et al. 2019). Knowing the genetic information can contribute to safe and effective sports training (such as preventing

sudden death, injury prevention, increasing the effectiveness of training programs), as well as better athletic recovery, medical care, and many other areas (Brazier et al. 2019). On the other hand, it should be noted that it must be supervised and not misused for unethical issues.

1.1.2 Single Nucleotide Polymorphisms

In addition to environmental factors in sports performance, especially after 1990, with the initiation of the Human Genome Project and revealing the importance of genetic information, the effect of genetic factors on sports success has gained great importance in recent years and various results have been revealed by researches (Ahmetov et al. 2009). The Human Genome Project is an international scientific research project for gene mapping, and it is known that there are approximately 88 million Single Nucleotide Polymorphisms (SNP) at the end of this project (Frazer et al. 2009).

While our genes show about 99.9% similarity among humans, 0.01% difference is differentiated by various variations. These variations consist of changes in SNPs that can be found inside or outside the protein-coding regions (Frazer et al. 2009). The frequency of presence of SNPs may vary within the population and between different geographic regions. SNPs are investigated with different study methods and tools such as case-control studies, genome-wide association studies and via generating knock-out model organisms. Not only in the genetic investigation of diseases, but also in some of our cognitive, physical, and behavioral characteristics can be explained by the presence of these variants.

When it comes to athletic performance, genes that affect many factors such as increased aerobic capacity, gene regions associated with strength and athletic ability, oxygen capacity, muscle structure, energy metabolism, and gene variants that cause these genes to be affected at different levels in each individual have been the subject of research for a long time. Increasingly, studies in the field of sports genetics show that genetic factors, as well as environmental factors, have important effects on sports performance. Many studies have revealed that genetic factors affect athletic performance depending on the type of sport (whether it's a team sport or individual sport). Although sports performance is also affected

by various factors, some single nucleotide polymorphisms have been associated with sports performance more than other tested genes (Ahmetov et al. 2015).

1.1.3 Methods in Genetic Testing

Especially associated with sports success; variations in single nucleotides in genes known to encode features such as muscle type and structure, muscle contraction rate, muscle elasticity, soft tissue resistance, healing and injury resistance, oxygen capacity, heart rate and blood pressure, lipid metabolism, glucose metabolism, exercise adaptation (SNP) detection and that it affects many parameters such as speed, strength and endurance in athletic performance, has been revealed by the researchers through genetic and physical tests on elite and non-elite athletes (Ehlert et al. 2013).

Two different DNA sources are used in companies that are sold commercially: saliva and blood (Goode et al. 2014). With these DNA sources, SNP research are analyzed differently; Chip array genotyping, Next generation sequencing (NGS) are most commonly used methods for comprehensive studies including sports genetics, nutrigenomics, risk of illnesses and general wellness insights (Goode et al. 2014).

Genome wide association studies (GWAS) are used for different sample types (I. Ahmetov et al. 2015). SNP genotyping methods with GWAS are used in oligonucleotide array, bead array, and micro arrays; However, these methods are expensive, not easily accessible and do not give results in a short time (Lambert et al. 2013; Patil et al. 2001). Another method, Chip array genotyping, is mostly used to understand complex organisms, to understand unstable population genetics, and to investigate evolutionary mechanisms (Lambert et al. 2013). Next Generation Sequencing (NGS), on the other hand, is quite expensive, although it creates a wider read on the genome in a short time. In SNP genotyping, Polymerase Chain Reaction (PCR) genotyping can be considered the best method in terms of the cheapest, easy access and speed (Collins 2010).

1.1.4 Investigation Methods Used to Find Associated Genetic Regions

Rapid sequencing of DNA has enabled the recognition of individual genetic variations that contribute to athletic performance with the development of technology (Gibson 2009). When all these types of genetic markers come together, there is a large genetic background link that influences these performance parameters. For this reason, researchers try to understand the genetic factors that affect injuries with these 4 study types; candidate gene studies, genome-wide association studies, case-control studies, and meta-analysis studies (Ryan-Moore et al. 2020).

1.1.5 Candidate Gene Studies

Scientists first examine the physical parameters of muscles, tendons, connective tissues, regulatory systems, or other related parameters to determine the factors they will examine, then investigate genetic factors that can affect these regions. For example, there are muscles, bone, connective tissues, and collagen; these phenotypes have been previously associated with more than 20 genotypes, including the COL family, which is responsible for collagen production. After obtaining these genes, studies look for Single Nucleotide Polymorphisms, which are the most common type of genetic variation among humans (Pruna et al. 2013 ; Gabbett et al. 2012). Even though most of the studies mention that there are more than 150 genes associated with athletic ability, replicated studies are not sufficient to make enough association studies among different populations (Larruskain et al. 2018; Gibson 2009).

In that case, single loci are not adequate to make precise assumptions in various sports branches and with different ethnical background of the athletes. For instance, at University of Sao Paulo, researchers have conducted a study with Brazilians to analyze the effects of sports relevant polymorphisms in a cohort of top-level athletes (Guilherme et al. 2018). Since the ethnical background of the Brazilian cohort was heterogeneous when compared to populations like Asians and Europeans, 23 different polymorphisms in 20 genes were

analyzed for association studies. Even some of the genotype distribution results were similar in Spanish endurance and power athletes' frequencies such as M268T polymorphism in the *AGT* gene, R577X polymorphism in *ACTN3* gene and C/T polymorphism in *NOS3* gene was similar to Spanish cohort. Conversely, G482S polymorphism in the *PPARGC1A* gene differed from endurance athletes from Spain, Russia and Poland (Guilherme et al. 2018).

To sum up, with the increase in such studies, it is necessary to search for new candidate genes, to test existing candidate genes in different populations, and to work with different types of athletes in order to understand the importance of the effectiveness of genes and the distribution of different alleles in the same gene (Guilherme et al. 2018).

1.1.6 Genome Wide Association Studies

Since the importance of testing various genes and alleles have been mentioned above, another crucial method to analyze and compare within the entire genome is Genome-Wide Association Studies, in which hundreds of thousands to millions of genetic variants are tested across the genomes of many individuals to identify genotype-phenotype relationships (Al-Khelaifi et al. 2019). Even if these methods are used for complex features such as diseases, they can also be determined in studies such as tendon injuries (Kim et al. 2017).

Genome-wide association studies (GWAS), in which hundreds of thousands to millions of genetic variants across the genomes of many individuals are tested to identify genotype-phenotype associations, have revolutionized the field of complex disease genetics over the past decade.

The fact that not every gene region has been studied in Turkish athlete cohorts is important in terms of finding new population and sport-specific variations by sequencing candidate genes that affect athletic performance. The study of genetic factors and new variations not found in different populations with similar approaches can provide de novo variants in similar genes that affect performance by comparing them with over 200 markers mentioned in the literature (Collins et al 2015; Gibson 2009). Although certain

polymorphisms have been studied in the Turkish population, there are no further studies on the discovery of new sport-specific gene variants yet (Eroglu et al. 2018; Ulucan et al.2015).

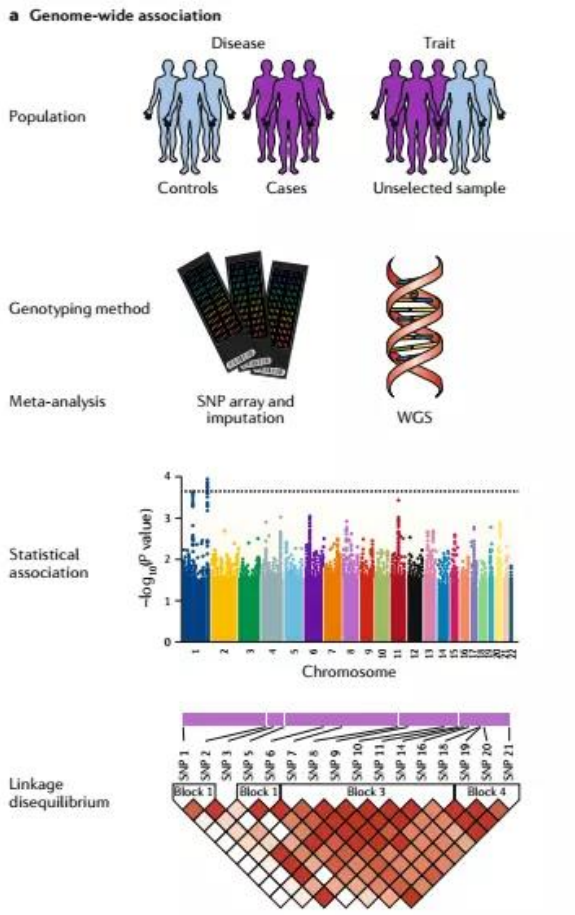

Figure 1.1: The structure of Genome-wide association studies (GWAS)

1.1.7 Case - Control Studies

To obtain which genes are associated with the phenotype, case-control studies are performed between athlete cohorts and controls; where genetic variants associated with athletic performance in the genomes of many individuals are tested to identify genotype-phenotype relationships (Ahmetov et al. 2015). Candidate genes were discovered after these

genes were studied in populations with specific athletes with muscle types, injury types or other specific physical parameters linked to athletic performance. For example, there are different SNPs associated for different subtypes of Hamstring injury: acute, overuse, severe, and repetitive (Larruskain et al. 2018).

Each injury subtype has a different associated SNP, so these studies are crucial to get the effect of which alleles show higher risk, statistically significant, and genetic as a risk factor. Even candidate genes are highly useful for studying known variants, but there may still be unidentified genes that may adversely affect soft tissue injuries in different populations. For instance, there are more than 30 genes associated with injury types among athletes (Larruskain et al. 2018; Baltazar-Martins et al. 2020; Collins et al. 2015). To obtain which genes are associated with the phenotype, case-control studies are performed between athlete cohorts and controls; where genetic variants associated with athletic performance in the genomes of many individuals are tested to identify genotype-phenotype relationships (Collins et al. 2015).

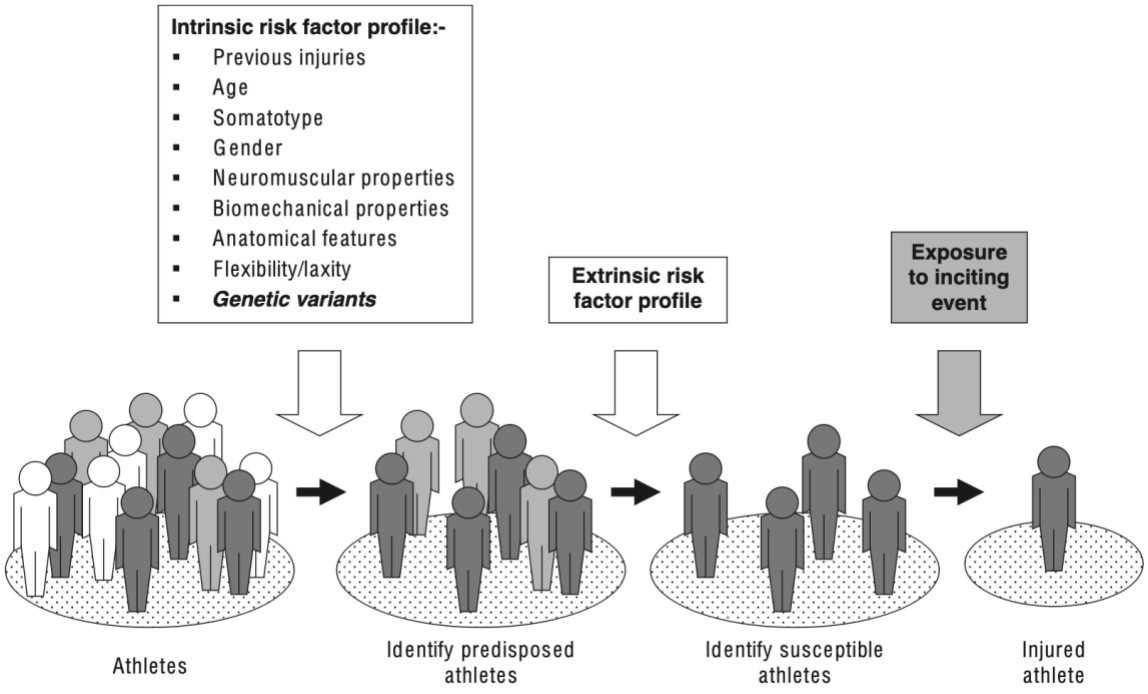

Figure 1.2: The working scheme used to explain the relationship between injured athletes and genes (Collins 2010)

Previously associated genes with injuries and muscle types are generally included in the studies. After examination of these genes with certain athletes with STI injury within populations, candidate genes are discovered.

1.1.8 Meta-analysis Studies

It is important to investigate the effect of gene regions in different populations on the phenotype and to find new interactions between genes according to injury types, with the increase of meta-analysis studies and case-control studies among these studies (Gabbett, Ullah, and Finch 2012).

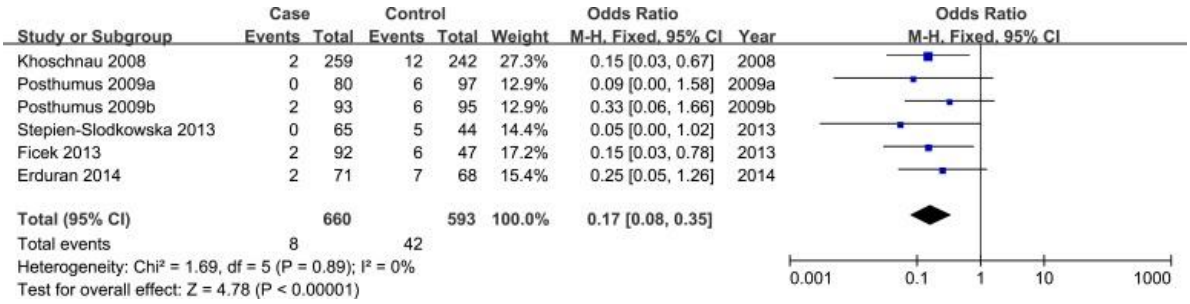

Figure 1.3: In meta-analysis, gene variants are determined regarding the alleles selected to investigate if there are any correlation within the studies to determine effects of the selected variants. After the papers are selected for the study, different models are tested to determine the overall effect (Tharabenjasin et al. 2019).

1.2 Previously Tested Genes

It is known to encode features such as muscle type and structure, muscle contraction rate, muscle elasticity, soft tissue resistance, recovery and injury resistance, oxygen capacity, heart rate and blood pressure, lipid metabolism, glucose metabolism, exercise adaptation, which are the main physical factors affecting performance in sports. Detection of variations

in single nucleotides (SNPs) in genes has helped to improve and investigate these physical effects of the performance parameters. It has been revealed by researchers through tests conducted on elite and non-elite athletes that these genes affect many parameters such as speed, strength and endurance in athletic performance (Ahmeto et al. 2009; Guth and Roth 2013). Although studies have been widely conducted on the detection of variations in gene regions such as *ACTN3* (alpha-Actinin-3), *ACE* (Angiotensin I Converting Enzyme), *PPARA* (Peroxisome Proliferator-Activated Receptor-alpha), it is thought that over 200 markers affect sports genetics (Eynon et al. 2013; Ahmetov et al. 2015). Understanding genetic determinants can clarify the criteria for physical activity, especially for athletes and individuals.

1.2.1 Sports Injuries

Long-term intense exercises, hours spent positively, and the appropriate genotype are sometimes not enough to raise a champion. Therefore, factors such as human biology, physiology, genetic research, and most importantly, environmental factors, rest, and loading should be carried out together. Sports clubs and federations lose more than $4 Billion a season to heal injuries that are ineffective and do not offer pretreatment (Opar et al. 2012).Team trainers and doctors are researching new exercise methods (such as Nordic hamstring exercise, isokinetic exercises) to overcome these problems (Engebretsen et al. 2008). Some gene variations have been determined to play a larger role in non-impact Hamstring injuries, and the expression analyzes of these genes have been performed and the protein levels after injury have been investigated in some articles (Collins 2010).

1.2.2 Soft Tissue Injury

Basically, soft tissue injury occurs with the damage in the muscles, ligaments, and tendons throughout the body. Different types of disability are also examined under the sub-title of soft tissue injuries (Collins et al. 2015). Injuries such as Achilles tendinopathy, tennis elbow, and tendon ruptures can be listed as the types of injuries most commonly encountered by athletes (Collins et al. 2015; Collins 2010). Soft tissue damage to muscles, ligaments and tendons are common in many sports activities like football, tennis, running, rugby etc. Most common symptoms that athletes encounter are the pain, swelling and bruising in the muscle (Hendersonn et al. 2010; Opar, et al. 2012). While these injuries are very important to health, when we think of the sports industry, injuries have a huge impact on the sports success and fitness of professional athletes.

Different types of disability are also examined under the subtitle of soft tissue injuries (Collins et al. 2015). Injuries such as Achilles tendinopathy, tennis elbow, and tendon ruptures can be listed as the types of injuries that occur under the sub-title of soft tissue injuries and which athletes encounter most (Feeley et al. 2008). When all these types of injuries come together, there is a large genetic background link that influences these injuries.

1.2.3 Common Soft Tissue Injuries

Amateur and professional athletes train according to their injury history after the problems they encounter or continue their training according to the physical/health tests carried out at regular intervals in the sports clubs they are affiliated with. However, such measures are insufficient to prevent new injuries in the long run and ensure that the athlete has a healthy season. Similarly, anthropometric tests used in athlete selection and systems that are insufficient to process the data are insufficient to predict the physiological characteristics of this individual in the future and the development of the athlete. Although injuries play an important role in sports performance, approximately two million people are treated each year for sports-related injuries (KOKU 2015) . Most injuries occur between the

ages of 13 and 18, resulting in failures among young athletes (Guth and Roth 2013).When the most common injuries among athletes were investigated, it was found that soft tissue injuries are frequently encountered in different fields in sports (Engebretsen et al. 2008). The most common soft tissue injuries can be listed as follow;

- Tendon Ruptures, partial or complete.
- Anterior Cruciate, medial or lateral collateral ligaments of the knee
- Rotator cuff tears of the shoulder
- Tennis Elbow
- Calf Strains
- Hamstring strains
- Quadriceps strains

1.2.4 Soft Tissue Injuries and Genetic Factors

Soft tissue injuries are very important for athletes; because, no matter what branch they are professional in, the most common cause of athletes quitting sports at an early age are recurring soft tissue injuries and subsequent losses (Pehlivan 2013). The main cause of these injuries are movements that do not require direct contact, such as sprains, excessive stretching/strain, tissue traumas that occur after injury, and sudden rotational movements. (Griffin et al. 2006)

From the previous studies, it was examined that there is strong evidence that some sports require different genetic backgrounds and so different physical requirements to be successful in different fields. Each injury subtype has a different associated SNP, so these studies are crucial to get the effect of which alleles show higher risk, statistically significant, and genetic as a risk factor (Eynon et al. 2013). Even candidate genes are highly useful for studying known variants, but there may still be unidentified genes that may adversely affect soft tissue injuries in different populations. Studies with cohorts of athletes from Europe have shown that *COL1A1* gene variants may affect the risk of injury (Eken 2018; C. Wang et al. 2017).

For candidate gene studies, genes and muscle types previously associated with injuries are often included in studies. Candidate genes were discovered after these genes were studied in populations with specific athletes with soft tissue injuries. For example, there are different SNPs associated for different subtypes of Hamstring injury: acute, overuse, severe, and repetitive (Larruskain et al. 2018). Each injury subtype has a different associated SNP, so these studies are crucial to getting which alleles show higher risk, statistically significant, and effect as a genetic risk factor. Even candidate genes are highly useful for studying known variants, but there may still be unidentified genes that may adversely affect soft tissue injuries in different populations (G. Wang et al. 2013). Since genetic frequencies in each population may be different, it is important to determine the frequency of gene regions and to investigate the effect of these variants on the phenotype and to find new interactions between genes according to injury types, with the increase of meta-analysis studies and case-control studies among these studies (Jones et al. 2017).

Although a genetic variation study has not yet been conducted in Turkey for specific soft tissue injuries such as Achilles tendinopathy or Hamstring injury, genes that cause soft tissue injuries abroad have also been examined in terms of Hamstring injuries, and it has been stated that some genes may have a positive effect on injuries (Eken 2018). Different types of disability are also examined under the sub-title of soft tissue injuries (Eroğlu 2015). Injuries such as Achilles tendinopathy, tennis elbow, and tendon ruptures can be listed as the types of injuries most commonly encountered by athletes (Collin et al. 2015).When all these types of injuries come together, there is a large genetic background link that influences these injuries.

1.2.5 Genes Associated with Soft Tissue Injuries

Inflammation after heavy training is a defense method used by the body to improve the immune system depending on genetic factors, to protect the body against environmental factors and to repair damaged tissue that may occur in the body after training. *COL5A1* is the gene involved in the production of collagen and is effective in determining the mobility of

muscles and tendons (Guo et al. 2022). *COL1A1,* involved in type-1 collagen synthesis, is found in tendons and muscle ligaments, so the function of this gene may determine injury rates (Pruna et al. 2013). *MCT1* and *IL-6* gene variants are involved in controlling inflammation after training (Opitz et al. 2015; Guilherme et al. 2018; Ben-Zaken et al. 2021). *GDF5* is an important growth factor involved in muscle recovery after sports. These gene variants constitute important factors that determine the speed of recovery of the athlete both after injury and possible injuries.

1.2.6 Soft tissue injury associated genes, variants, and their effect on the phenotype

The list of associated genes and alleles in different papers have shown that there are more than 50 gene regions specified in diverse sub-types of injuries. Table 1.1 summarizes associated gene regions previously related with soft tissue injuries.

Table 1.1: Gene names, SNP ID's (rs numbers) and phenotypic effect of the genes are described from the previously studied soft tissue injury variants.

Genes	SNP	Effect On Phenotype
CRP	rs1205 (C/T)	It is involved in various functions related to host defense due to its ability to recognize damaged cells and initiate their elimination in the blood (Kasapis and Thompson 2005).
IL-6	rs1800795 (G/C)	*IL-6* is a pleiotropic cytokine expressed in immune and muscle cells. It is involved in a wide variety of biological functions, including regulation of differentiation, proliferation and survival of target cells (Reihmane and Dela 2014).
GDF5	rs143383 (G/A)	Growth Differentiation Factor 5 regulates the development of cell types involved in numerous connective and bone tissues, including cartilage, joints, teeth (Stastny et al. 2019).

COL1A1	rs1800012 (C/A)	It encodes the pro-α1 chain in the structure of type I collagen. Type I is a fibril-forming collagen found in most connective tissues and abundant in bone and tendon (C. Wang et al. 2017).
COL3A1	rs1800255 (G/A)	Type III encodes the pro-a1 chain. It is generally found in flexible/extendable connective tissues (O'Connell et al. 2015).
COL5A1	rs12722	It encodes the pro-a1 chain found in the structure of type V collagen (Stastny et al. 2019).
MCT1	rs1049434 (T/A)	The transport protein catalyzes the rapid transport of many mono carboxylates such as lactate and pyruvate across the plasma membrane (Fedotovskaya et al. 2014).
MMP3	rs679620 (C/T)	Matrix Metalloproteinase 3 is the enzyme involved in the degradation of fibronectin, laminin, collagen III, IV, X and IX and cartilage proteoglycans (Gibbon et al. 2017).
TNC	rs2104772 (T/A)	*TNC* is the intercellular matrix protein, responsible for neural regeneration and synaptic plasticity (Lulinska-Kuklik et al. 2019).
TIMP2	rs4789932 (A/G)	*TIMP2* metalloproteinase inhibitors are natural inhibitors of matrix metalloproteinases, a group of peptidases involved in the degradation of the extracellular matrix (Lulinska-Kuklik et al. 2019).

1.2.7 The Collagen Type-1 Alpha Gene *(COL1A1)*

COL1A1 is one of the genes that can determine injury risk rates because it helps to synthesize collagen (Type-1). Type-1 collagen is found in the tendons and ligaments of the muscle, so its functions can determine injury rates (Eken 2018). *COL1A1* is located on chromosome 17 with its 17q21.33 location, and mutations in this gene cause various congenital (Mendelian origin) diseases such as Caffey's Disease, osteogenesis imperfecta type I-IV and acute soft tissue ruptures (Eroğlu 2015). Interestingly, however, other variations in this gene that do not have a major effect have been suggested to have effects on

soft tissue injuries in athletes. For example, those with the AA genotype in the rs1800012 SNP have a lower risk of soft tissue injury and are more protective against soft tissue injuries than the CC phenotype. The rs1800012 SNP is a functional SNP, and the AA allele in this SNP may cause an increase in *COL1A1* expression, which has been suggested to result in the formation of stronger tendons and ligaments (C. Wang et al. 2017).

Among soft tissue injuries involving different sports branches such as tennis, football, skiing, long and short distance runners ; Although there are more than 50 studies on the rs1800012 variant, which has also been examined with subheadings such as Achilles tendinopathy, anterior cruciate ligaments, tennis elbow and Hamstring injuries, the rs1800012 variant is not the only variant in the *COL1A1* gene (Pruna et al. 2013; Zhong et al. 2017). Studies that focus on only one or a few candidate SNPs lead to ignoring the SNPs that are around the candidate SNP and have the main impact. Our bioinformatic analysis with genome data obtained from world populations indicate that 22 different variants located near the rs1800012 variant may also be important in soft tissue injury Table 2. To reveal these variants, Sanger sequencing can be performed to investigate whether novel variants are associated with the selected football cohort in this thesis. Thus, the effects of the variants to be found on soft tissue injury will be examined and variants specific to the Turkish population will be revealed. In addition, it is important in terms of finding new interactions between genes according to injury types (Gabbett et al. 2012).

Table 1.2: A list of other variants found in the vicinity of the rs1800012 variant for the *COL1A1* gene, with the variant type and locations listed

Variant ID	Location	Molecular consequences
rs1268471442	50,200,365	intron variant
rs1907979004	50,200,366	intron variant
rs1470437670	50,200,367	intron variant
rs2696252	50,200,373	intron variant
rs1907979694	50,200,376	intron variant
rs1038490257	50,200,380	intron variant
rs1907980051	50,200,384	intron variant
rs376722154	50,200,385	intron variant
rs1907980465	50,200,386	intron variant
rs1179960317	50,200,387	intron variant
rs1800012	50,200,388	intron variant
rs1193822382	50,200,389	intron variant
rs959444558	50,200,393	intron variant
rs540903804	50,200,395	intron variant
rs1907981785	50,200,396	intron variant
rs1907981975	50,200,398	intron variant
rs2696253	50,200,402	intron variant
rs1025256775	50,200,403	intron variant
rs1598302682	50,200,404	intron variant
rs1323782357	50,200,405	intron variant
rs1282739467	50,200,408	intron variant
rs966614411	50,200,409	intron variant

1.2.8 The alpha Actinin 3 (*ACTN3*)

ACTN3 is the gene region that encodes the alpha-actinin-3 protein in humans. It is mostly produced in type II fast-twitch muscle fibers. Type II muscle fibers are responsible for rapid and forceful muscle contraction, but they are less resistant to faster muscle fatigue and unusual muscle injuries. This parameter plays an important role in choosing the type of training in terms of maintaining the muscle structure and energy of the individual (Baltazar-Martins et al. 2020).

It is known that the change in the ACTN3 gene affects athletic status, adaptation to training, risk of injury, and the way muscle fatigue is seen. Many studies with elite athletes have revealed that the *ACTN3* R577X variant is effective in injuries and training adaptation (Clos et al. 2019). It has been stated that the easily injured structure of muscle fibers may be more vulnerable to injury with the higher rate of alpha-actinin-2 protein produced in the absence of alpha-actinin-3 protein. An increase in ankle injuries is also observed in individuals with the X allele (Appel et al. 2021). It has also been revealed that individuals with the R allele have higher adaptation to eccentric exercises. Also in another study, individuals with the XX genotype presented with higher serum creatin kinase levels and muscle pain values than RR individuals after a protocol of eccentric knee extensions (Vincent et al. 2007). Similarly, XX soccer players showed higher serum creatine kinase concentrations than their R allele counterparts after a session of plyometric exercise (Pimenta et al. 2021). The athletes carrying the X allele presented higher reductions in jump height, and higher values of serum creatine kinase and self-reported muscle pain than RR athletes after a marathon or a half-ironman triathlon (Del Coso et al. 2017).

Perhaps, α-actinin-3 plays a role during the eccentric phase of endurance exercise activities that confers a higher capacity to the muscle, to resist muscle damage despite the restricted expression of this protein to fast-twitch fibers. It has been reported that the variants of XX within soccer players (first league division, Italy) had almost a three and twofold higher probability of suffering muscle injuries in general and severe muscle injuries, respectively, than their RR counterparts (Del Coso et al. 2019).

1.3 Muscle Types

The training type is important in determining whether the athlete's muscle type has explosive power or the endurance characteristic that responds better to prolonged intense training. There are two muscle types in our body, Type 1 and Type 2; Type 1 muscle fibers make up slow-twitch muscles, while type 2 muscle fibers make up fast-twitch muscles (Pickering and Kiely 2017).

Table 1.3: Associated genes for muscle types and their effects on phenotype from previous case-control studies (cont. on next page)

Genes	SNP	EFFECT ON PHENOTYPE
ACTN3	rs1815739 (C/T)	Stabilizes muscle contraction in fast-twitch muscle fibers (type II) in the presence of alpha-actinin-3 (Baltazar-Martins et al. 2020).
PPARa	rs4253778 (G/C)	It regulates liver, heart and skeletal muscle lipid metabolism, glucose homeostasis, mitochondrial biogenesis, cardiac hypertrophy (Lopez-Leon et al. 2016).
PPARGC1a	rs8192678 (C/T)	It regulates fatty acid oxidation, glucose utilization, mitochondrial biogenesis, thermogenesis, angiogenesis, formation of muscle fibers (Gineviciene et al. 2016).
AGT	rs699 (G/A)	Angiotensinogen is an essential component of the renin-angiotensin system, which regulates vascular resistance and sodium homeostasis, thereby determining blood pressure (Guilherme et al. 2018).
ACE	rs4646994 (I/D(490/190 bp))	It regulates circulatory homeostasis by synthesis of vasoconstrictor angiotensin II and degradation of vasodilator kinins (Cieszczyk et al. 2009).

BDKRB2	I/D(-9/+9)	The bradykinin receptor is responsible for modulation of vascularization, inflammation, edema, pain, cell proliferation, muscle contraction, and glucose metabolism (Neto et al. 2020).

While the maximum size that Type-1 muscle fibers can reach is limited, Type 2 muscle fibers are the type of muscle that reaches the highest speed and strength among all muscle groups. Type 2 muscle fibers respond well to strength and strength training and tend to grow more than other types of muscle. For example, in the studies performed with weightlifting athletes, because they have fast-twitch muscle fibers, they gain a larger appearance at the end of their training, while marathon runners develop their thinner and longer muscles because they train slow-twitch muscle fibers (Del Coso et al. 2019).

1.3.1 Tested Genes for Muscle Types

1.3.1.1 The a-actinin-3 gene (ACTN3)

ACTN-3, which is considered the most important gene of sports performance, is involved in the formation of actin fibers, which play an important role in the functioning of our muscles. Thanks to the differences in the proteins encoding this gene, some changes may occur in the structure of our muscles. This gene provides athletic performance, adaptation to exercise type, and recovery from injuries. ACTN3 genotype can be used to indicate whether you will be an endurance or strength athlete, as well as how you will respond to different types of training (Eynon et al. 2013; Ahmetov et al. 2015).

Although there are 3 different genotypes for ACTN-3, people with the RR and RX alleles do not have any problems in the production of actin. Individuals with the XX allele are not affected by α-actinin-3 deficiency; because the ACTN-2 gene is activated and they compensate for the ACTN-3 deficiency (Ben-Zaken et al. 2015; Gineviciene et al. 2011).

However, as muscle fiber composition will change, their responses to different sports and training programs may also change. ACTN3 XX genotype averages were expressed at 25% in Asians, 18% in Caucasians, 11% in Ethiopians, and 3% in US African Americans and only 1% in Kenyans (Pickering and Kiely 2017). Some studies have stated that the XX genotype is observed more frequently in countries with low temperatures and that the X genotype may play a role in the evolutionary process (Houweling et al. 2017).

If the player's result is RR, the rate of producing type II muscle fibers, which is a fast twitch muscle group, is high because it has the R allele. It is more adaptable to the type of training based on explosive power. Muscle fatigue can be seen more. May be more resistant to training-induced injuries than Type I muscle types. The flexibility of the muscles may be lower (Broos et al. 2016; Baltazar-Martins et al. 2020).

If the player's result is RX, there is a high chance of adapting to a certain amount of explosive power training and endurance training. Because they carry the X allele, they are more likely to experience muscle injuries and ankle injuries. Rapid muscle fatigue may occur due to carrying the R allele (Broos et al. 2016; Baltazar-Martins et al. 2020).

If the player's result is XX, which is a carrier of only the X allele, it is highly adaptable to endurance training. Because they have a higher percentage of type I muscle fibers, the flexibility of the muscles and their resistance to fatigue are higher than the carriers of the R allele. There is a possibility of injury in muscle injuries and exocentric exercises (Broos et al. 2016; Baltazar-Martins et al. 2020).

1.3.1.2 The Angiotensinogen Converting Enzyme Gene *(ACE)*

Angiotensin converting enzyme (angiotensin converting enzyme), produced by the ACE gene, acts as a transforming protein in the RAS system, which has an important effect in regulating blood pressure. While inhibiting growth hormone causes the protein to work less actively, it provides reabsorption of salts and fluids, which play an important role in regulating blood pressure, in the kidneys (Fleming 2006).

A variation in the ACE gene affects the amount of production of the ACE protein. ACE gene may also have a polymorphism or genetic expression change that has been shown to

express itself in exercise performance. ACE gene expression can have two changes: insertion (insertion) or deletion (deletion). Studies have revealed the effects that individuals with the insertion (I) allele have a lower amount of ACE expression, and that type I muscle type is produced at a higher rate. It is stated that individuals with the deletion (D) allele respond better to strength-based training, and the type IIx muscle type has higher mass (Kang et al. 2012).

In individuals with II genotype, it causes a decrease in ACE activity, allowing players to improve their performance in endurance-like modalities. It has been stated that individuals with the I allele may be associated with an increase in type I muscle fiber size (slow twitch muscle fiber); that is, the presence of the I allele increases the mechanical efficiency of the skeletal muscle. Accordingly, your athletes with this allele can increase their adaptation to endurance sports at a higher rate by increasing endurance exercises in training. The insertion genotype (II) maintains energy balance during intense and prolonged physical exercise (Williams et al. 2000; Puthucheary et al. 2011; Ahmetov et al. 2015).

If the player is ID, due to the fact that he carries both alleles, the intensity that the player will give to strength or endurance training in the type of training he will choose may increase his adaptation to this training type. Individuals with the ID allele can have both types of muscle; therefore, you can choose training types for your player that require both explosive power and endurance (Williams et al. 2000; Puthucheary et al. 2011; Ahmetov et al. 2015).

Individuals with the D allele (D) have higher circulating ACE activity at higher levels in serum and tissue. It has been revealed that individuals with the D allele may have a high proportion of type-II (fast-twitch) muscle fibers. Accordingly, the training program based on explosive power increases the adaptation of the player to such exercises at a higher rate (Valdivieso et al. 2017).

1.4 How Multiple Genes Affect Each Other?

Sports genetics is associated with a multivariate heredity and examining more than one parameter related to sports performance in studies is important in terms of associating the results with each other. With the introduction of personalized medicine into our lives, genetic tests have become widespread, and we have more information about the genome. With the investments made in genetic tests, the quality of human health is improving day by day (Gibson 2009; Jones et al. 2016).

Examination of different variations on the same gene plays an important role in generating linkage maps of different loci on the same gene in genetic linkage studies to be performed in bioinformatic analysis. Methods and study cohorts have not been studied in the Turkish population before; so it is a pioneering study in the interpretation of the outputs and the database to be created in large-scale studies. These outputs can be further investigated for different sport branches in the Turkish cohorts.

1.5 Total Genetics Score and Sports Genetics

The complex structure of the genome has been provided with more comprehensive explanations thanks to analyses with bioinformatics tools. Genetic score calculations, which are used in bioinformatics and biostatistics, are used as a part of personalized medicine in clinical studies today and it has been stated that they can be an important tool in revealing important metabolic diseases such as obesity risk, Type-2 diabetes, cardiovascular diseases, heart attack by scoring on an individual basis (Leonska-Duniec et al. 2018; Collins 2010; Fawcett and Barroso 2010). These results, revealed by different tools and statistical approaches, reveal the importance of genetic risk score (Szulkin et al. 2015).

Elite athlete status is expressed in studies in which more than one gene has complex characteristics with gene-gene and gene-environment interactions. In the literature, although meta-analyses express the effect of each gene on endurance or strength-based athletic performance individually, they are insufficient to explain the effect of these genes together.

It reveals the effect of more than one marker on the endurance or strength profile, and the total genetic score calculation (TGS), which has been frequently used in recent years, is effective. Total genetic score (TGS) is expressed in the literature as a simple additive model that expresses the integrative effect of multiple markers on a trait (de la Iglesia et al. 2020). TGS not only determines the athletic profile in sports performance; it is also an approach used to estimate the risk of diseases such as Type-2 Diabetes and cancer that significantly affect public health on an individual basis (Leonska-Duniec et al. 2018).

By using TGS, physical performance increase, VO2max capacity and even nutritional tendencies according to the profile of the athletes according to the TGS score were also revealed. In the scoring studies on sports genetics, variants were tested with a different scoring system and a patent study was conducted for the genetic combination of physical tests according to the development of those athletes with the measurements obtained after the strength and training program. In the follow-up, while the athletes with the endurance profile according to TGS showed improvement in low-intensity resistance training; Athletes with a strength profile made improvements in high-intensity resistance training (Feito et al. 2018).

In the same study, no significant improvement was observed when athletes with the endurance profile according to the TGS performed high-intensity training for 8 weeks. Similar results were observed when athletes with a power profile completed a low-intensity training period (Feito et al. 2018). As a result of the study, the possible mechanism by which the polygenic profile of the athletes (the profile formed as a result of genotyping consisting of 15 polymorphisms) is associated with the training responses, the connection between genetic variations and skeletal muscle characteristics such as muscle type was established. Statistically significant results were shown in the results of the counter movement jump (CMJ) and aerobic 3-minute cycle test (Aero3), which measure endurance and explosive power in athletes who were included in the training program and trained in accordance with the genetic profile. In addition, 5 of the 15 gene polymorphisms included in the algorithm within the scope of the tests (*ACE* I/D, *ACTN3* rs1815739 C/T, *PPARA* rs4253778 G/C, *PPARGC1A* rs8192678 G/A and *VEGF-A* rs2010963 G/C) were determined by muscle fiber type. As a result of these studies, it sets an example that an algorithm can be developed and patented (Pickering et al. 2018; Feito et al. 2018).

Common sports genetics tests test for individual alleles and report non-cumulative results, but this result is insufficient. In this algorithm, when the relationship of each variant with each other is made numerically meaningful, the tendency of the individual to the training type is better understood. These results reveal the effect of performance increase by applying the total genetic score to the training program in revealing the genetic profile of the athlete (Kujala et al. 2020).

Physical parameters affecting sports performance have a polygenic (multi-genetic) structure controlled by more than one gene; It is not sufficient to regulate the training of athletes by looking at a single gene allele and genotype and to base it on athlete selection criteria. Therefore, the development of a quantitative scoring system, considering the genotype of more than one gene, has been the preferred method in recent years. In this approach, called the total genetic score (TGS), the number of risk alleles of the genes whose effect on this phenotype has been proven in the literature in order to determine the susceptibility of an athlete to a positive or negative phenotype in terms of sports performance is obtained by linearly summing (Pranckeviciene et al. 2021).

1.6 Physical Parameters Used for Physical Tests and Infrastructure

Since physical exercise is a factor that contributes to the quality of life of people, it is recommended as a very important prescription for both health and the therapy of diseases. In addition, the type of exercise, intensity, competitions, traumas and stress cause both physiological and metabolic changes in the human body (Ntanasis-Stathopoulos et al. 2013).Therefore, the relationship between athletic performance, heavy exercise and physiological changes should be well known. Although athletes are defined as physically normal and healthy, psychophysiological stress arising from intense training, competitions and matches can change their homeostasis, biochemical and hematological results (Pareja-Galeano et al. 2014; Vina et al. 2012).

Mobility is one of the key metrics of sports performance. The flexibility capacity of muscles and tendons plays an active role in many branches, especially in football. The range

of motion (range of motion, ROM) of the muscles and joints gives us important information about the injuries that the athlete may experience during the training and the maximum performance he can get during the training period (Siegrist 2008).

Although the benefits of regular exercise and the response of the human body to physical activity are well known in recent years, their genetic background is unknown (Ahmetov et al. 2015). Understanding genetic determinants may clarify the criteria for physical activity, especially for athletes. From this point of view, collagen types that make up the structure of muscles, joints, bones, and tendons provide important information about the structure of the skeletal and muscular systems we have. Collagen proteins, which have structurally different subunits and isoforms, are found in different tissues, in various proportions, and contribute to the stiffness or flexibility of cartilage, connective tissue, muscles and bones at different rates (Wilson et al. 2019; Maffulli et al. 2013).

In case control studies to discover genetic predispositions in players, injury data of players are required for statistical analysis. At this point, the physical parameters to be obtained from the players, their injury history, the duration of their stay in the game and their recovery period can provide the necessary data for the association studies. In this case, some biochemical analysis, injury history and devices that measure physical parameters can be effective in determining people's strength and risk of disability. Although the frequency of measuring the data, the number of repetitions, the time of obtaining the data after the match are variable factors, association studies can be made using different biostatistical analyzes and long-term follow-ups can help to provide preventive data for sports injuries (Lulinska-Kuklik et al. 2019; G. Wang et al. 2016; Gibbon et al. 2017).

One of the devices that can be used to test Hamstring injuries and other soft tissue injury determinants is the Nordbord device, which gathers data for isometric and eccentric hamstring strength. Although the range of the data obtained varies due to different factors such as the player's height and weight; the power difference between the two legs, the leg muscle strength at different angles, the power obtained after the injury can be determined thanks to the Nordbord device, and a relationship can be established between players' injuries according to these measurements (Hogberg et al. 2022; Buchheit et al. 2016).

1.7 Hypothesis And Specific Aims of The Thesis

In the literature, even though genetic association studies have been carried out, the difference between the sports branches and the physical association studies performed differ from each other. At this point, lack of genetic information in Turkish athlete cohorts may result in inaccurate genetic test outputs.

Some of the shortcomings in the studies performed are that previously determined variants are not tested in every population. So, mutations between populations or new variants that can be found at different frequencies are not included in the common qPCR tests, which may create a gap in the discovery of new variants. Gene variants associated with soft tissue injuries are seen not only in exon regions but also in non-coded regions, so examining different variations on the same gene plays an important role in drawing link maps of different loci on the same gene in genetic linkage studies to be performed in bioinformatics analysis.

The first hypothesis of the thesis is *COL1A1* genetic variants affect the soft tissue/hamstring injury risk in professional soccer players. Moreover, there are other variants in *COL1A1* in addition to rs1800012 that potentially affect the soft tissue/hamstring injury risk. These variants will be discovered by Sanger sequencing the rs1800012 nearby gene region.

As there are more than 20 genes associated with soft tissue injury, the second hypothesis is that besides the *COL1A1* gene, the *ACTN3* rs1815739 (C/T) and the *ACE* rs4646994 (I/D(490/190 bp)) variants will influence soft tissue injury risk in the professional soccer cohort.

Finally, the power profiles of the players were compared with the average data from 1000 genomes by calculating the total genetic scoring method of the tested *ACTN3, COL1A1* and *ACE* genes. Since the athletes are at the elite level, our hypothesis suggests a higher scoring result from the European population.

In line with these hypotheses, the aims and objectives of the project are as follows:

To reveal variants in *COL1A1* gene reported in the literature to be associated with soft tissue injury, and other variants in the immediate vicinity of these variants, in samples taken from professional football players by DNA sequencing.

To determine the minor and reference allele frequencies of both previously reported, and the potential novel variants to be found in the study cohort/ and as a pioneering study for the first time examine their effect on hamstring injury in the Turkish population by statistical analyses.

To conduct statistical association tests between the discovered genetic variants and soft tissue injury.

Testing for the first time the total genetic score approach to model genetic risks for power predisposition in Turkish soccer players.

CHAPTER 2

MATERIALS AND METHODS

2.1 Determination of Athletes to Participate in the Study and Collection of Samples

A cohort study design was adopted, and participant recruitment was performed. 21 volunteer professional football players over the age of 18 are included in the study. The study cohort consisted of licensed athletes with similar physical profile characteristics and with the same nutrition, resting and training schedules. Those who do not comply with these criteria were not included in the study. 3cc blood samples were taken into EDTA tubes from each volunteer participating in the study was taken by authorized club doctors. Interview forms were filled by authorized club doctors and coaches in line with the opinions of the athletes.

Permission was obtained from the ethics committee of Izmir Democracy University for the study. In addition, support letters were received from İzmir Göztepe football club.

A questionnaire form was prepared to be directed to Göztepe Sports Club football players, one of our Super League football teams, consisting of questions such as the age of the athlete, weekly training hours, number of competitions, rest time, and injuries suffered, as well as their physical characteristics. Athletes filled out the questionnaires and the results were evaluated statistically. In addition, in order to measure the hamstring strength of the athletes, by using the Nordbord machine through club physiotherapists, the right and left leg strengths at 90 degrees were taken in Newtons and added to the data set. Although the applied muscle strength is directly proportional to weight and height, lower strength values were observed in athletes who had hamstring injuries.

Following collection of blood samples, physical data to be collected regarding soft tissue injury:

- Injury histories of the athletes in previous seasons,

- Injury frequency, injury types and rest days during the 2020-2021 season, and recovery time
- Nordbord data received during the 2020-2021 Season
- Groin Bar, Thermal camera, Lactate scale 4 data
- Height
- Weight
- How many years they have been involved in football
- Weekly training hour
- Competition time
- History of soft tissue injury
- Soft tissue injury occurrence
- Age ranges when the soft tissue injury occurred

2.2 Candidate Gene DNA Sequencing Studies

2.2.1 DNA isolation protocol

DNA isolation from blood samples were done with the "ROCHE High Pure PCR Template Preparation" kit, following the protocol of the kit as;

The blood sample was taken from the EDTA tubes and put into the 1,5ml Eppendorf tubes. Then 200µl tissue lysis buffer and 40µl Proteinase K was added and mixed and incubated for 1 hour at 55℃. After the incubation 200µl tissue lysis buffer was added and incubated for another 10 minutes at 70℃. Then 100µl of isopropanol was added and mixed well.

After these steps, the sample was transformed into a High Pure Filter tube within the collection tube. Then the samples were centrifuged for 1 minute at 8000xg. After each centrifugation, the flow was discarded and put into a new collection tube. 500µl of Wash Buffer was added into the filter and centrifuged again for 1 minute at 8000xg. This process

was repeated one more time with a 500µl Wash Buffer and centrifuged. Then to remove the excess residuals, the tubes were centrifuged at top speed for 10 seconds and the filter tube was put into a new sterile microcentrifuge tube. For the collection of DNA, 30µl prewarmed Elution Buffer was added into the filter and centrifuged at 8000xg for 1minute. This step is repeated two times to get more DNA products and then the solution was taken to the Nanodrop for the analysis of the DNA purity.

The concentrations of the isolated DNAs will be measured by spectrophotometric methods using Nanodrop. The concentration and purity of the DNA samples obtained will be determined by measuring their absorbance at 260 and 280 nm wavelengths. It is expected that the A260/A280 absorbance ratio of the ideal purity quality DNA will be between 1.8-2.0 values. Isolated DNA samples were stored at -20°C until sequencing.

2.2.2 Targeting and sequencing of candidate SNPs in the candidate gene

2.2.2.1 *COL1A1* Genotyping

The human *COL1A1* gene reference sequence was downloaded from the National Center for Biotechnology Information (NCBI, https://www.ncbi.nlm.nih.gov/) and Ensembl Genome Browser (https://www.ensembl.org/index.html) databases. and Primer3 program (http://bioinfo.ut.ee/primer3/) primers were designed to perform DNA sequencing 400 bases up and 400 bases downstream of the SNPs targeted in these genes.

That is, genetic sequencing of both candidate SNPs and the genomic environment of these SNPs will be performed. Primers targeting these sequences were designed with Primer 3 (http://bioinfo.ut.ee/primer3/), IDT (https://eu.idtdna.com/pages), NCBI, UCSC Genome Browser databases and programs. With the help of designed primers, polymerase chain reaction (PCR) will be carried out by Thermal Cycler device using Taq polymerase and dNTP set in the samples. At the end of PCR analysis, agarose gel electrophoresis will be performed to observe band formation in PCR samples as;

Table 2.1:Polymerase Chain Reaction volumes per sample for *COL1A1*

Component	Reaction Volume per Sample
Forward Primer	1.5 µl
Reverse Primer	1.5 µl
dNTP	1 µl
PCR Buffer	5 µl
Taq Poymerase	0.2 µl
DNA template	3 µl
Water	37.8 µl

Table 2.2: PCR conditions for COL1A1 gene

PCR Steps	Time	Reaction Degree
Initial	3 minutes	94℃
Denaturation	30 seconds	94℃
Annealing	30 seconds	59℃
Extension	1 minute	72℃
Number of cycles		35x

After the band formation was observed in the agarose gel, Sanger Sequence analysis was performed from the samples whose expected PCR product size is seen as a single band at the DNA marker level on the gel (ABI 3130XL Sanger Sequence device, IZTECH BIYOMER).

2.2.2.2 *ACE* Genotyping

For *ACE* rs4644994 variant, designed primers were ordered. With the help of designed primers, polymerase chain reaction (PCR) was carried out by Thermal Cycler device using Taq polymerase and dNTP set in the samples. At the end of PCR analysis, agarose gel electrophoresis was performed to observe band formation in PCR.

Table 2.3: Polymerase Chain Reaction volumes per sample for *ACE*

Component	Reaction Volume per Sample
Forward Primer	1.5 µl
Reverse Primer	1.5 µl
dNTP	1 µl
PCR Buffer	5 µl
Taq Poymerase	0.2 µl
DNA template	3 µl
Water	37.8 µl

Table 2.4: PCR conditions for *ACE* gene

PCR Steps	Time	Reaction Degree
Initial	3 minutes	94℃
Denaturation	30 seconds	94℃
Annealing	30 seconds	59℃
Extension	1 minute	72℃
Number of cycles		35x

1 gram of agarose is put into 100 ml of 1x TAE buffer and dissolved by boiling. After preparing the solution, it is left to cool down to 60 degrees and 7.5ml EtBr gel is added to the solution and poured immediately after mixing.

2.3 Bioinformatics analyses and annotation of genetic variants

After the Sanger sequence was made, outputs were taken and analyzed first by UGENE. The sequencing format used in UGENE was ".ab1" for each sample. To search for the selected variant of rs1800012, location was determined from the NCBI database. Each sample was aligned, and peaks were observed by scanning the sequence, for the variant, G or T allele should be observed for the polymorphism. So, when only the G allele is observed, the sample is known to be Homozygous GG, and when only T is observed it is known to be Homozygous TT. When both G and T allele had been observed, then it is known to be GT, which is heterozygous.

Then to observe new variants in general for Multiple Sequence Alignment, data was uploaded to the MEGA7 program in the format of ".masx" and whole samples were aligned

to observe the new variants. To compare the new variations, a complete sequence was retrieved from the NCBI database. Since Sanger sequencing generally starts to give the peak for detecting the nucleotides after 200 base pairs, weaker signals might generate wrong outputs to determine the exact sequence. After making the alignment, these nucleotide changes were noted, and their locations were determined for the comparison of the original sequence.

2.4 Statistical Analysis

After analyzing each sample of the athletes, characteristics of the population, polymorphism frequencies were calculated for the respective variant. Variants of 3 different gene regions selected in the studied football cohort were genotyped and the frequencies within the cohort were determined.

After determining the polymorphisms for 3 different gene variants, all data was gathered in Excel with the physical parameters and survey results to make the statistical analysis. Each parameter was aligned, and all missing data was changed to NA for statistical analysis. Then for determining averages, standard deviations, and other population statistics, the SPSS Statistics program was used. For the simple statistics, descriptive statistics were done from the program and characteristics of the football players were determined. For the t-test and ANOVA analysis, R Studio Program was used.

2.4. Calculation of Total Genetic Score

The aim of the calculation of the Total Genetic Score is to give a quantitative value to the predisposition of the athlete to this quality by attributing the numerical values of zero (0), one (1) and two (2) depending on the frequencies of the genotypes associated with a

determined attribute in the literature. The following formula was used for the calculation of total genetic score;

TGS = (100/2n) (G1 + ... + Gn)

The variable n in the formula represents the number of genes required to calculate the parameter, and the variable G refers to the genotypes scored.

Table 2.5: Power scores for *ACTN3, ACE* and *COL1A1* genes, their ID, position, and allelic variation

Gene	RS ID	Chromosome	Position	Alleles	Allelles in User Interface	Power Score
ACTN3	**rs1815739**	**11**	**66560624**	**TT**	**XX**	**0**
				CT/TC	**RX**	**1**
				CC	**RR**	**2**
ACE	**rs4644994**	**17**	**61565900**	**DD**	**DD**	**2**
				ID/DI	**ID**	**1**
				II	**II**	**0**
COL1A1	**rs1800012**	**17**	**50200388**	**GG**	**CC**	**0**
				GT/TG	**CA/AC**	**1**
				TT	**AA**	**2**

2.5 Meta-analysis

For meta-analysis, peer-reviewed English, and Turkish published genetic association studies of *COL1A1* variants with soft tissue injury are gathered. Allele and genotype frequencies, case-control numbers are recorded. Allelic model, additive model, dominant model, and recessive model meta-analyses were performed by the 'meta' package in the R environment.

CHAPTER 3

RESULTS AND DISCUSSION

As a result of the questionnaire made with the athletes, 21 athletes aged between 18 and 34 participated in the study. The effort tests to measure athletes are similar in terms of transparency of the test process. It is important for the precision of the results to be obtained at equal time intervals. According to the results obtained, it is recommended that the physical tests and recordings in the training programs to be similar. Having at least half of the athletes tested is very important in maximizing the confidence interval.

In the study, the study was conducted with the senior A Team consisting of athletes with an average of 18 years of experience, whose professional playing years ranged from 10 to 27 years. Although the average age of these athletes is 25.8, it consists of a team that trains for an average of 11 hours a week throughout the season. Among these athletes, 7 out of 21 athletes had Hamstring injuries, 5 out of 7 athletes had Hamstring injuries once and 2 of them more than once. Also 15 of the athletes had at least one soft tissue injuries out of 21 athletes. The average height of the players is 183.38cm, with an average BMI of 23.59. More detailed physical parameters, minimum and maximum values and standard deviations are shown in the table.

In interviews with both Göztepe Football Club, it has been reported that there are at least 10 soft tissue injuries each season. Therefore, even if the number of athletes to be included in the project is limited to 20 volunteers due to the budget of the project, soft tissue injury data will be available in sufficient numbers to make statistically meaningful analyzes. At the same time, since the historical injury data of these athletes will be examined, it is expected that the number of injury data will increase even more.

Table 3.1: Characteristics of football players.

Variable	Total N=21
Age (years) (mean±SD)	25.81 ± 4.7
Years played (mean±SD)	18 ± 4.5
Height	183.38 ± 8.2
Weight	79.53 ± 9.36
BMI (mean±SD)	23.59 ± 1.53
Training hours	11.14 ± 0.9
Hamstring injury	
Yes, N(%)	7 (33%)
No, N(%)	14 (67%)
Other soft tissue injury	
Yes, N(%)	15 (71%)
No, N(%)	6 (29%)
Nordport (Right leg)	381.38 ± 77.00
Nordport (Left leg)	373.52 ± 78.55

3.1 Targeted primer sequence

From the targeted primer sequence, the designed primer was set as;

- **Forward primer;** CCACCTCTAGATCTGGAAAGTAAAG
- **Reverse primer;** GAAAGAGTTACAGCCTCCCTG

In this primer sequence, it was aimed to observe the *COL1A1* variant of rs1800012 and possible variants located in the 800 bp target zone. Previously investigated locations were determined from the NCBI database regarding the previous studies. After the Sanger sequencing, it is aimed to investigate the variants in Figure 3.2.2 is observed in selected athletes.

3.2 PCR Results for Primer Design

PCR analysis was performed to observe the working of the targeted primers. At the end of the analysis, 1% agarose gel electrophoresis was performed to observe band formation in PCR samples. From the results, the band formation was observed so the designed primer sequence worked properly.

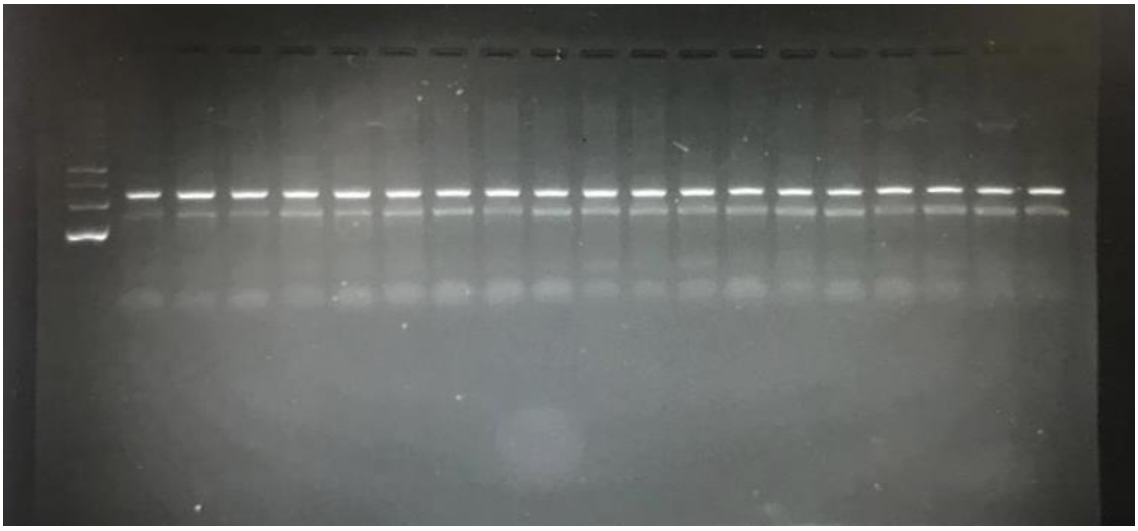

Figure 3.1: Images of 1% Agarose gel obtained as a result of PCR analysis performed on selected athletes

3.3 Sanger Sequence Results

After testing the primers to work, the samples were sent to IZTECH BIOMER for sequencing. The DNA sequence chromatograms to be obtained after the sequence analysis were examined in the UGENE (http://ugene.net/) program and the rs1800012 variant was determined in each athlete (Figure 3.2).

3.3.1 UGENE Outputs

The sequences obtained for each sample after the Sanger sequence were analyzed in the UGENE program, and the single nucleotide polymorphisms within the 1000 base region were examined. In the image, rs1800012 heterozygous G/T sample was examined the 354th base pair of the sequence. For each sequence of the athletes, variants of G allele and T allele is obtained. From the total of 21 samples, all sequences were analyzed and checked twice with the reverse sequence.

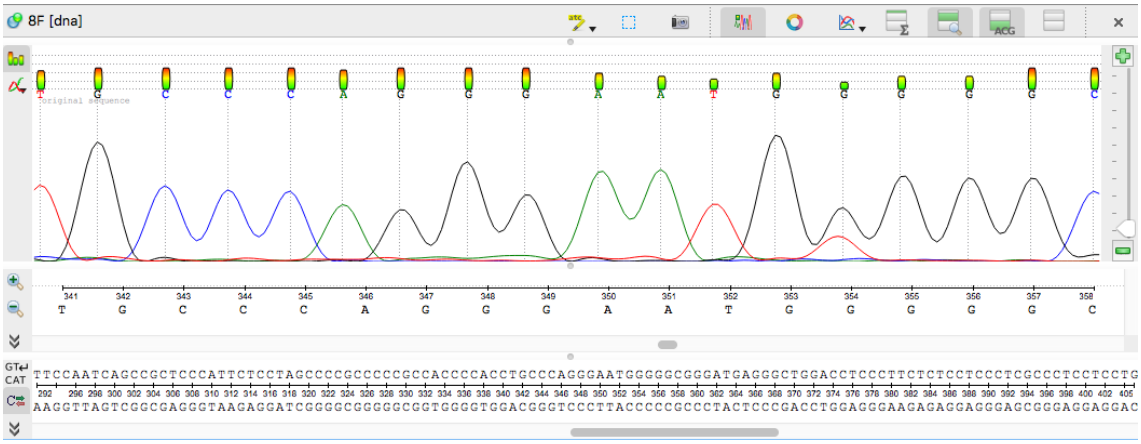

Figure 3.2: Target sequence was analyzed with UGENE Program to detect variants within selected region.

3.3.2 MEGA Outputs

From the outputs of Multiple Sequence Alignment, new polymorphisms were spotted and quality of the whole sequences were checked from UGENE if there are *de novo* polymorphisms to be investigated from the NCBI database. Some results due to low signals resulting from sequencing within the 1000 base region were examined again in the UGENE program, and locations of new variants different from the population with different polymorphisms were noted.

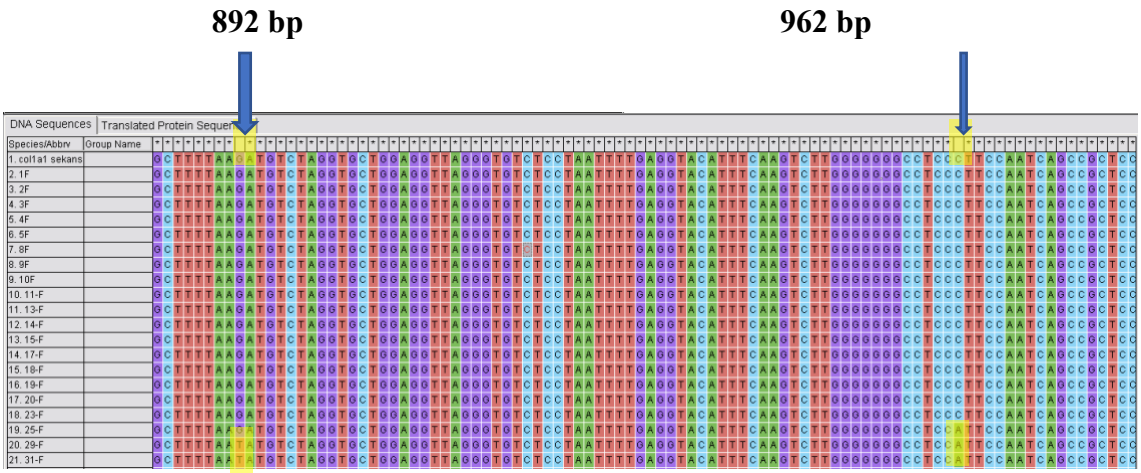

Figure 3.3: UGENE results of the samples were analyzed with multiple sequence alignment to check for new variants.

3.3.3 Multiple Sequence Alignment

For bioinformatics analyses, all sequences were combined using the MEGA Program (Figure 3.3.3-1). Multiple sequence alignment was performed using the Clustal algorithm and aligned to the *COL1A1* gene sequence found in NCBI.

The gene regions, whose locations were determined by comparison with the reference sequence, were identified via NCBI for comparison with variants already identified in different studies. From the results, the sequence alignment results were not significant to examine the role of specific variants to be found in this study in soft tissue injuries in addition to the variants reported in the literature.

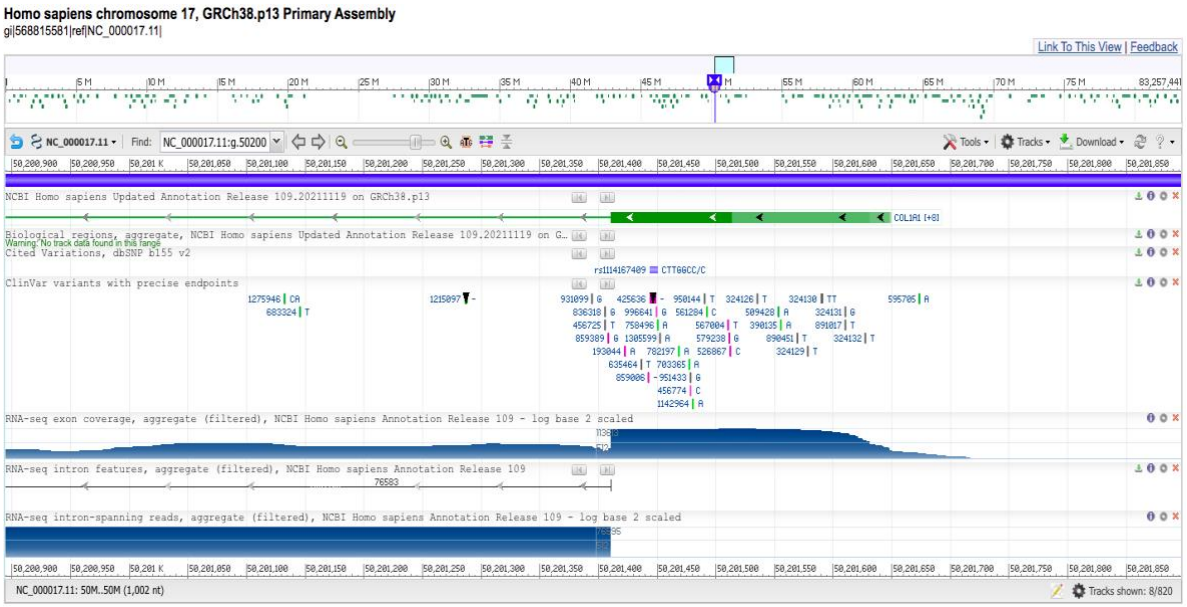

Figure 3.4: The locations of other variants in the *COL1A1* gene with other polymorphisms in the existing population.

From the Sanger sequence results, after the samples were sorted, the topmost sequence was aligned to be the NCBI primary sequence and comparisons were made between samples to detect polymorphisms at different locations by performing multiple sequence alignment. After checking each sample with the Sanger sequence for the new polymorphisms, there was no significant polymorphism since all the detected polymorphisms were in the end of the sequence so the reading errors with few signals may have occurred.

3.4 Interpretation of Allele and Genotype Frequencies of Genetic Variants

After analyzing each sample of the athletes, polymorphism frequencies were calculated for the respective variant. Variants of 3 different gene regions selected in the studied football cohort were genotyped and the frequencies within the cohort were determined. When compared with the allelic European frequencies from 1000 Genomes database, *COL1A1*

rs1800012 G/T polymorphism showed exact similarity with the European population statistics.

According to the results shown in the Table 10, for *COL1A1* gene rs1800012 variant; GG genotype was higher than GT genotype in football players. The TT genotype was not observed in the population. The frequency of the G allele was higher than the T allele when the allele distribution was made. Although TT genotype was not observed in the studies; It was determined that 14 of 21 athletes had GG genotype and 7 athletes had GT genotype. Athletes with the GG genotype may have a higher risk of soft tissue injuries because their COL1A1 metabolic rate is slower. These results show that if athletes are injured, there may be a potential genetic risk of injury.

For the *ACTN3* rs1815739 variant, frequency of the European population was also similar to the genotyped cohort. R (C) allele in the European population was found 56% while, in the cohort the population frequency was 52%. Since distribution is nearly equal in both alleles, football players require both endurance and power traits, so the results and distributions were significant when compared to other studies. When alleles are examined, the fact that the RX allele is the most abundant allele in the population may contribute to the orientation of the footballers to the strength or endurance profile depending on the synthesis of alpha actinin and to turn into type 1 or type 2 muscles depending on the training type.

For the *ACE* rs4644994 variant, frequency of the European population was different than the genotyped cohort. I allele in the European population was found 50% while, in our cohort the population frequency was 27.5%. The main cause is that only 1 athlete has the II allele and most of the players were either ID or DD variant. This means moderate or increased *ACE* synthesis, which means that athlete profiles have a high propensity for power.

Table 3.2: Distribution of *COL1A1, ACE, ACTN3* variant allele frequencies and genotypes in the soccer cohort

Gene (variant)	N (%)	Frequency in Europe
COL1A1 (rs1800012)		
G	37 (84%)	0.84
T	7 (16%)	0.16
Genotypes		
GG	14 (67%)	
GT	7 (33%)	
TT	0	
ACTN3 (rs1815739)		
R	22 (52%)	0.56
X	20 (48%)	0.44
Genotypes		
RR	2 (9.5%)	
RX	18 (85.7%)	
XX	1 (4.8%)	
ACE (rs4644994)		
I	11 (27.5%)	0.50
D	29 (72.5%)	0.50
Genotypes		
II	1 (0.05%)	
ID	9 (45%)	
DD	10(%50)	

3.5 Population Statistics

In the cohort, average age for Hamstring injury is 28 years. Hamstring injury was expected to increase with increasing age. In the analyzes performed, statistically, there is a statistically significant difference between the age of people without hamstring injury and the age of those with injury ($p<0.05$).

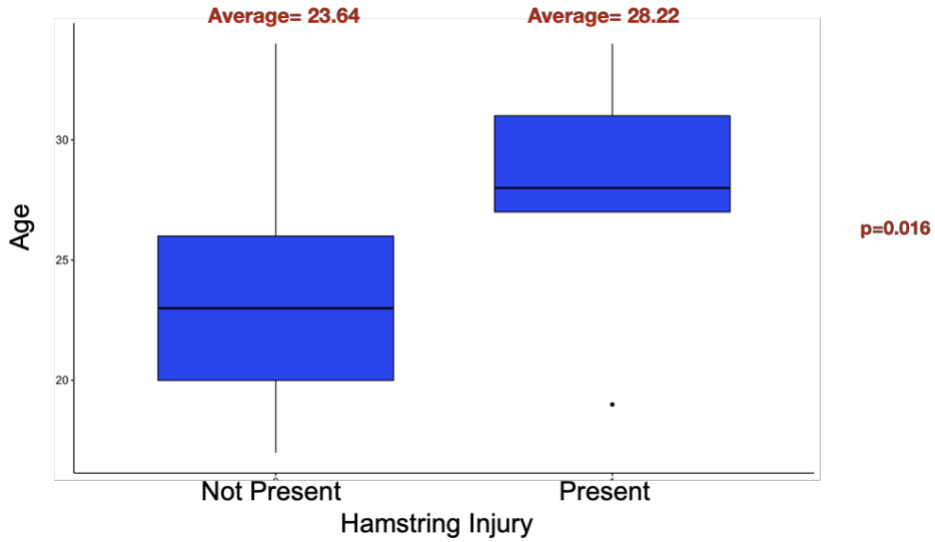

Figure 3.5: Hamstring injury occurrence and age of football players.

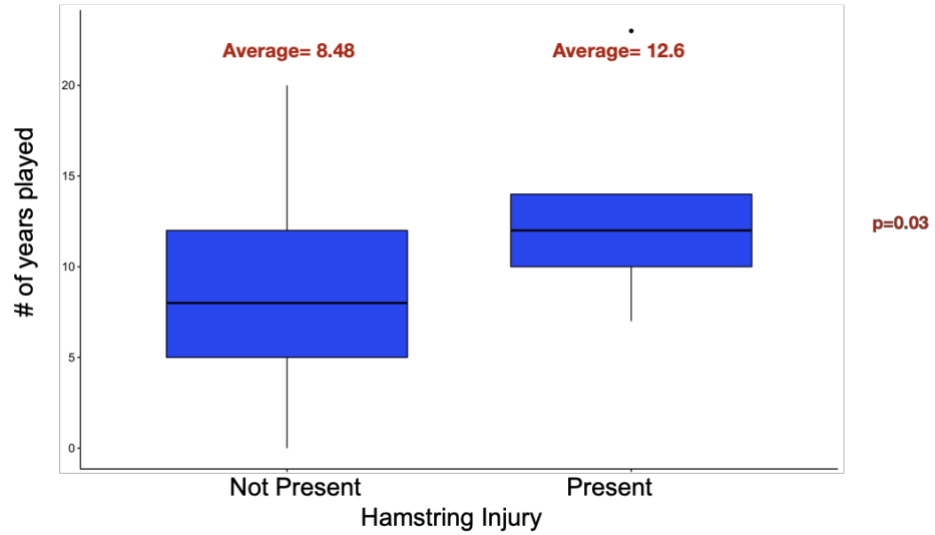

Figure 3.6: Hamstring injury occurrence and number of years played of football players.

As the previous analysis, there is also relationship between the years of playing as a professional athlete and Hamstring injury presence in the cohort ($p<0.05$). This result was also expected since the training hours increase, the possibility of getting injured also increases. As the fatigue after training, consecutive seasons, and injuries due to age increase, it is expected that the injuries of the players will also increase.

3.6 Statistical Analysis for Injury Types and *COL1A1* gene

In the statistical analysis made with athletes' injury history and *COL1A1* genotype results, no significance was found with Hamstring injury and *COL1A1* gene. From the previous studies, T allele has a protective effect since it is increasing the Type-1 collagen. Besides that, there was a slight significance with soft tissue injuries and *COL1A1* gene. From the results, it was seen that the players with T allele had less soft tissue injuries compared to the GG genotype.

From the other statistical analyses, there was no correlation between Nordbord results and *COL1A1* gene, though the average Nordbord measurements were 371.28 N (right leg) and 259.35 N (left leg) for GG allele and 401.57 N (right leg) and 401.85 N (left leg). Since the sample size is too low, even there is a significant difference between GG allele and GT alleles in both legs, the results were not statistically significant. In the Nordbord analysis, the measurement results were also discussed.

In the analyzes made to measure the relationship between the frequency of playing of the players and the injury genes, analyzes were made considering the number of matches played per week. As a result, a relationship with the number of matches and the *COL1A1* gene was not observed.

Table 3.3: Statistical Analysis for *COL1A1* rs1800012

	COL1A1		
Variable	**GG (14)**	**GT (7)**	**P value**
Soft tissue injury			
Yes	10 (72%)	2 (29%)	0.06*
No	4 (28%)	5 (71%)	
Matches played per week			
0	2	0	0.57
1	5	3	
2	7	4	
Nordbord (Right leg)	371.28 ± 76.6 N	401.57 ± 79.6 N	0.42
Nordbord (Left leg)	359.35 ± 80.6 N	401.85 ± 71.4 N	0.24

To determine the effects of different allelic models, another statistical analysis were done to select the best model for *COL1A1* gene. Since there were no TT alleles in the cohort, both genotypic and G-recessive models showed exact similarity. G-dominant model was also insufficient since there was no TT allele in the cohort. From these results shown in the Table 3.6-1, there was no association between *COL1A1* gene and Hamstring injury, though there might be an association between soft tissue injuries in general with *COL1A1* gene.

Table 3.4: Association of *COL1A1* rs1800012 genotypes with Hamstring and other injuries

Model	Hamstring Injury			Other Soft Tissue Injuries		
Genotypic	**0 (N=7)**	**1 (N=14)**	**P**	**0 (N=12)**	**1 (N=9)**	**P**
GG	9	5	0.74	10	4	0.06
GT	5	2		2	5	*
TT	0	0		0	0	
G-dominant						
GG/GT	14	7	-	12	9	
TT	0	0		0	0	
G-recessive						
GG	9	5	0.74	10	4	0.06
GT/TT	5	2		2	5	*

3.7 Nordbord Results

To determine the other association studies between Nordbord results and *ACTN3*, *ACE* and *COL1A1* gene, statistical analysis was done with the R Program to check if there is a correlation between genes and power status of the athletes.

From the results, there was no association between Nordbord results and genotypes. Although Nordbord is a device that helps athletes to establish an injury relationship because it measures Hamstring injury, the reason for the lack of a significance between genes may be due to the small number of samples, only one measurement is taken from the athletes during season, or the difference in weight and power averages of the athletes.

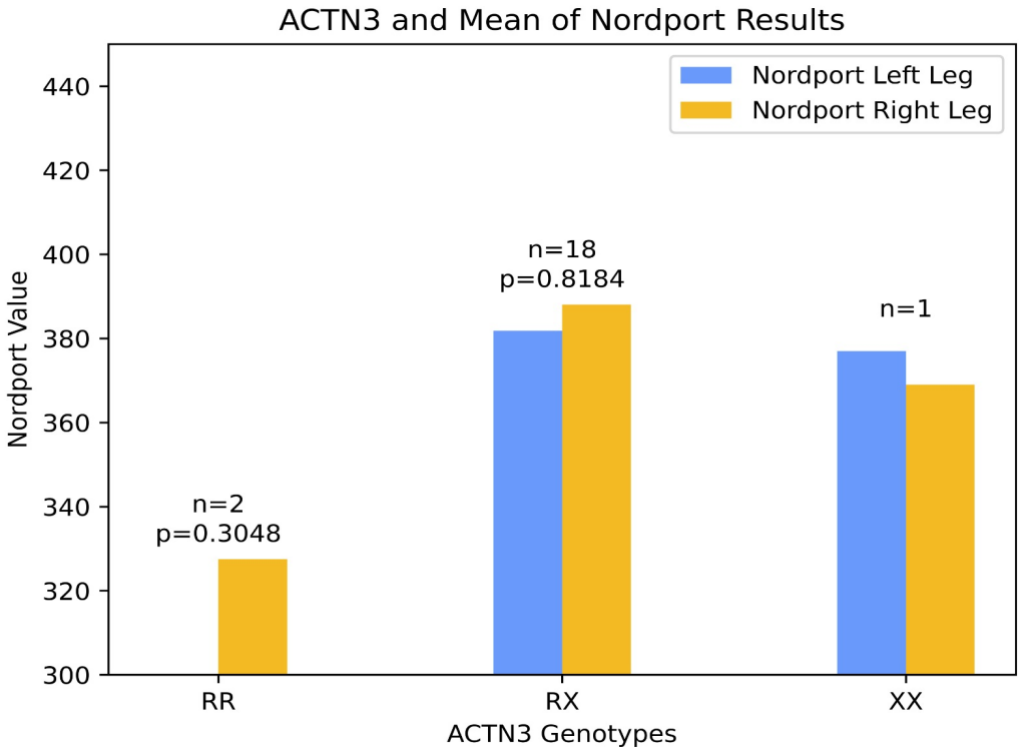

Figure 3.7: Distribution of *ACTN3* in the soccer player cohort

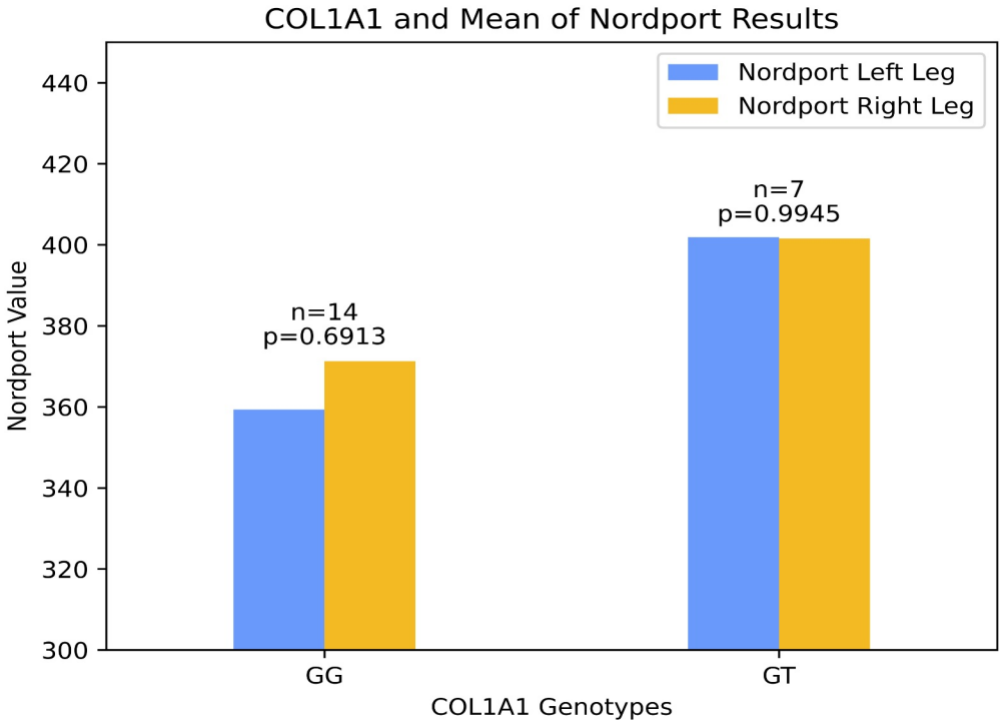

Figure 3.8: Distribution of *COL1A1* in the soccer player cohort

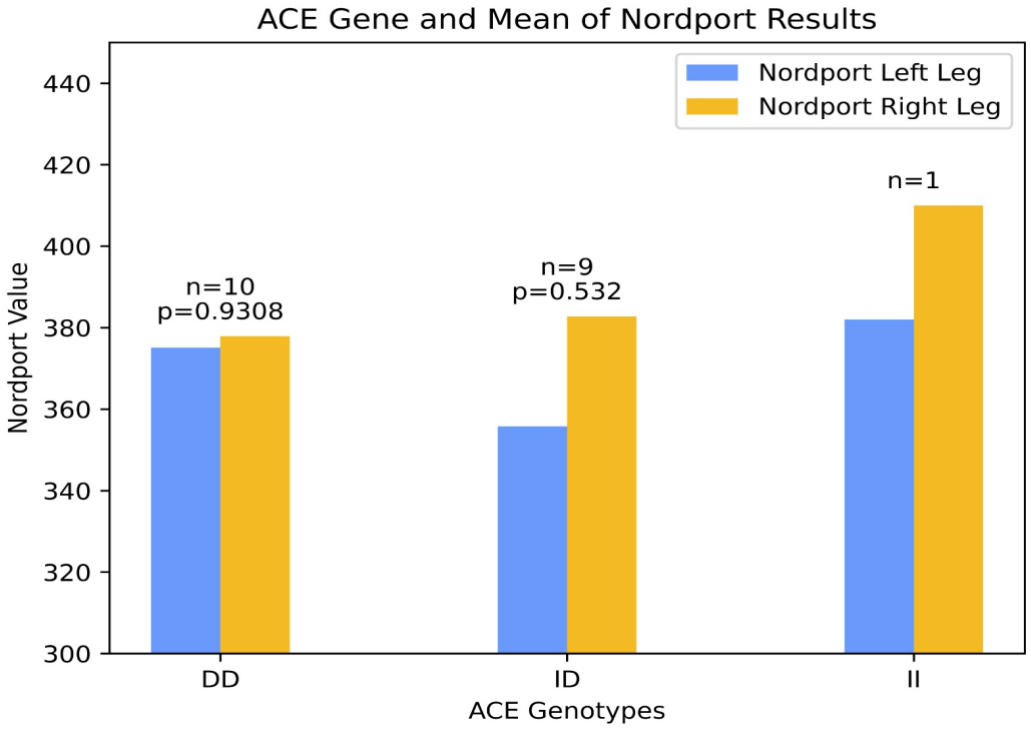

Figure 3.9: Distribution of *ACE* in the soccer player cohort

3.8 Meta-analysis results

From the previous studies done with the *COL1A1* gene and soft tissue injuries, studies for meta-analysis were collected and genotype data in studies associated with the *COL1A1* gene of soft tissue injury were extracted. For the analysis, different models in meta package were done to investigate the effect, odd ratio and statistically meaning models for effect size of the study. Since the variance is estimated based on the sample size, each study has various effect size and using different models with similar ethnic backgrounds of samples could be beneficial for the effect of the gene variant in this study.

The importance of the project is to obtain genes previously associated with soft tissue injuries and to test for intergenic interactions in the Turkish cohort. Even if there are positive correlations regarding genes, data are obtained from various populations such as Caucasians or Polish athletes, and due to ethnic differences, these genes may not be equally distributed in the Turkish population, so there may be a lack of genetic information in Turkish athlete

cohorts. From the meta-analysis results, the cohort showed similarity with the previous studies made with other populations.

In the meta-analyses, the additive and allelic model among 4 different models were expected to be similar to other studies. Since the T allele is also included in the analysis in other models, having no athletes with the TT allele in our analysis, it can be considered normal to see a difference. From the results, for the G allele odd ratio was found as 3.34 in dominant model and 3.54 in additive model.

The results of the meta-analyses for soft tissue injuries and the rs1800012 variant in the *COL1A1* gene were similar to other studies, summarizing that the effect of the G allele was also associated with soft tissue injury.

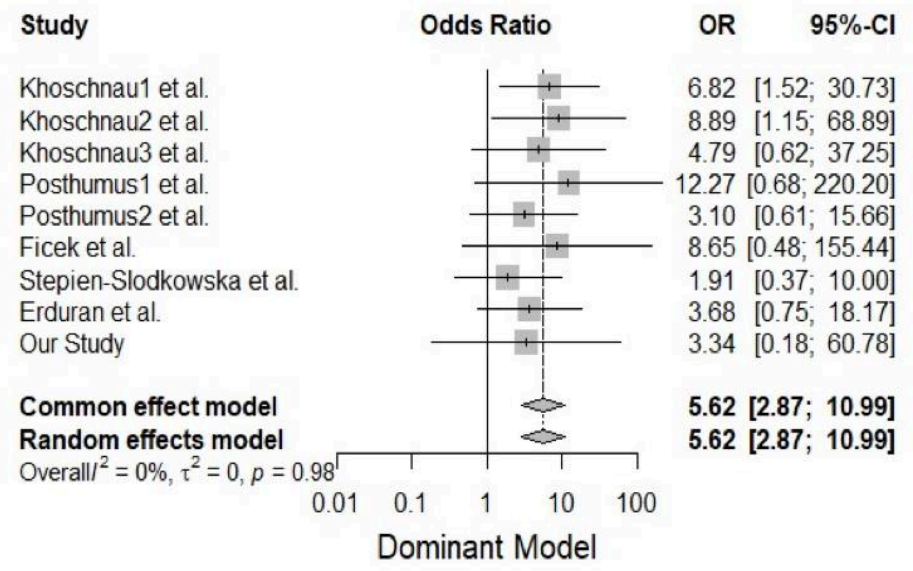

Figure 3.10: Dominant model compares GG versus GT + TT

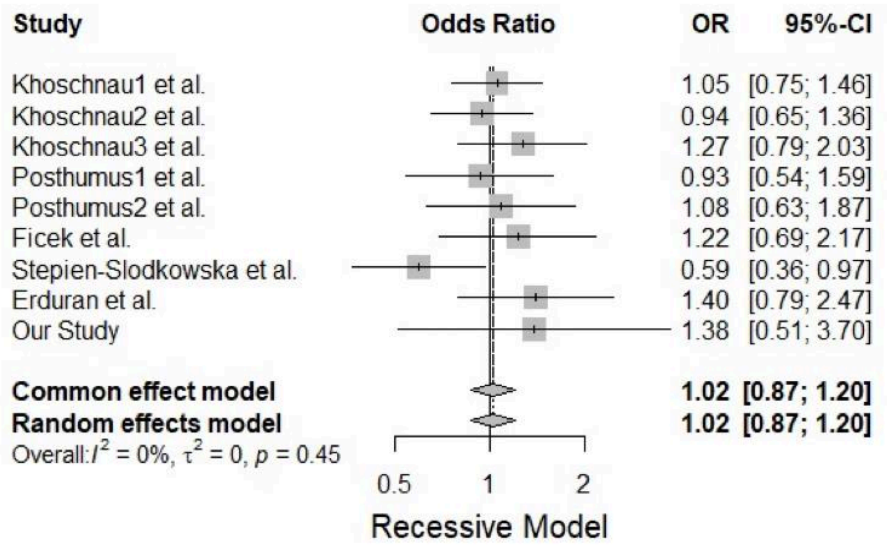

Figure 3.11: Recessive model compares GG + GT versus TT

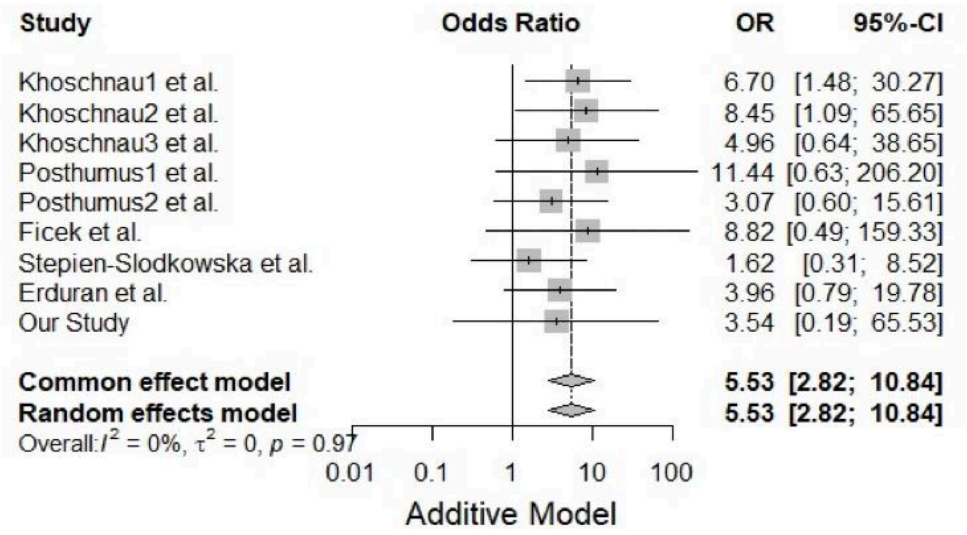

Figure 3.12: Additive model compares GG versus TT

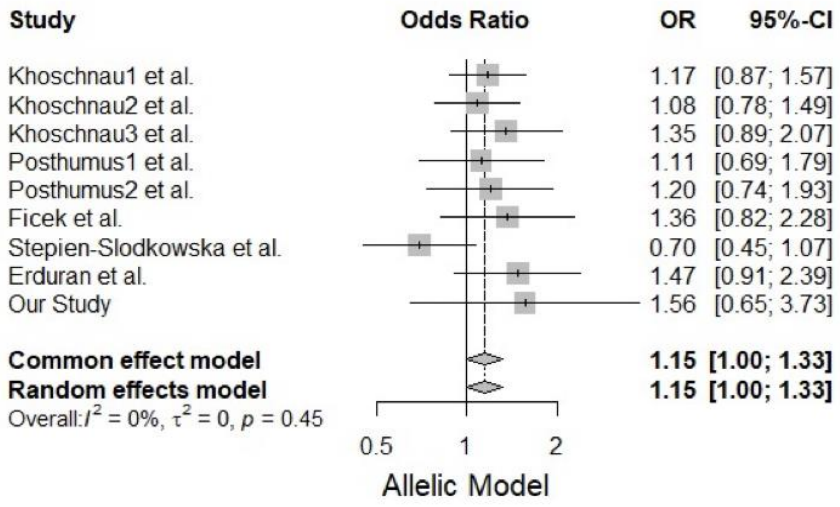

Figure 3.13: Allelic model compares G versus T

3.9 Total Genetic Score Calculations

Since sports performance has a complex structure, genetic variations among genes have an influence in athletic performance. Due to the polygenic traits in our genes, more comprehensive methods could be used to determine the correlation between genes and performance of the athletes. For the genetic profiling and scoring of the athletes, total genotype score (TGS) was proposed to evaluate the effects of the variants in a broader assessment. By calculating the Total Genetic score, the effect of each gene on the determined parameters can be calculated quantitatively based on the sum of alleles.

Gene-gene interactions of genes might be an effect of other pathways which is also related with collagen production, muscle type and injury risk. There are several studies that imply that the effect of the gene-gene and pathway interactions may also play a role in the performance of the athletes.

From the results, the aim was to reveal the statistical relevance of the genes by scoring 3 genes and the most studied variations, then evaluate the statistical analysis with physical

parameters derived from performance analysis in the season and evaluate if there is correlation with injury types and selected polymorphisms in the selected cohort.

Total genetic score of the calculated cohort was higher than European region, which shows higher Power predisposition in the football players. It is expected since the NCBI database consists of sedentary samples without any specific physical background. For the athletes' power scores, the median score was calculated and compared with each player in the cohort. 13 of the athletes showed higher power profile and mostly all the athletes' power scores were more than the general European scores.

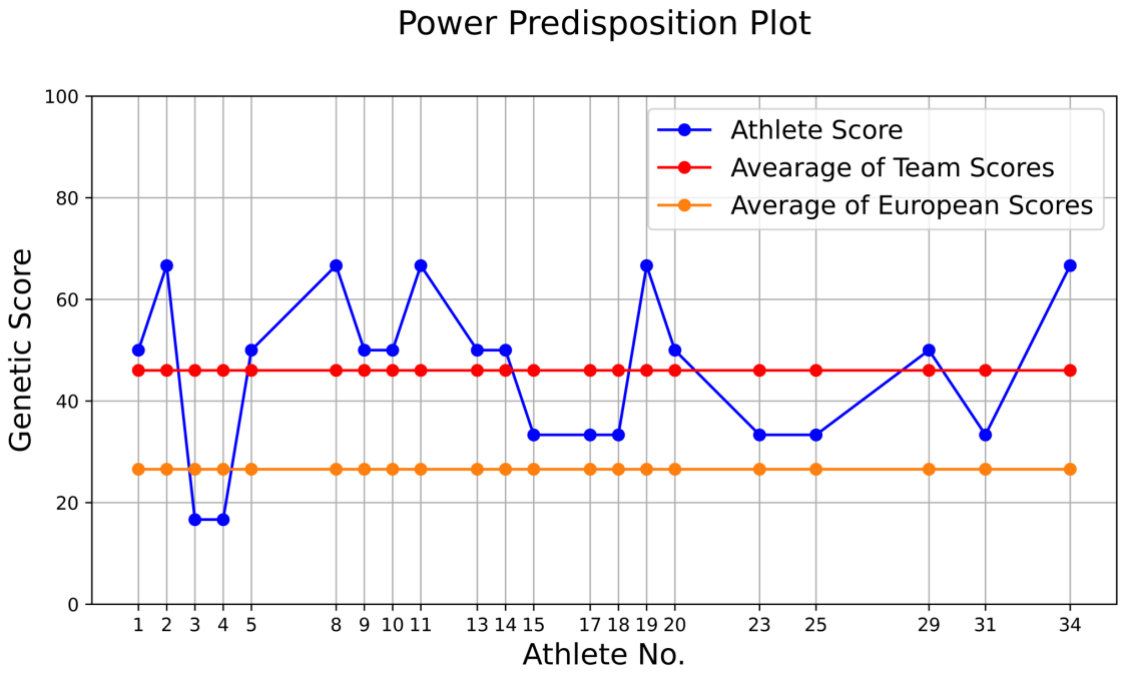

Figure 3.14: Calculation of Total Genetic Score based on the cohort and European population

CHAPTER 4

CONCLUSION

Soft tissue injuries compromise individual performance and team success in many sports. While injuries play an important role in sports performance, approximately two million people are treated each year for sports-related injuries. As mentioned, most injuries occur between the ages of 13 and 18, resulting in failures among young athletes (KOKU 2015). Main objection in this study is to investigate elite athletes' both genetic backgrounds with soft tissue injuries, which are multifactorial, by comparing the injury types and occurrence rates they have experienced with the associated physical data. Another aim of the thesis is to determine the regions that can be associated with new gene variants and to examine the relationship between the genetic structures of the athletes by making a biostatistical evaluation among the gene regions that have already been tested.

In soft tissue injuries, although there are more than 50 studies on the rs1800012 variant, which has also been examined with subheadings such as Achilles tendinopathia, anterior cruciate ligaments, tennis elbow and Hamstring injuries; it is not the only variant in the *COL1A1* gene. Each injury subtype has a different associated SNP, so these studies are crucial to get the effect of which alleles show higher risk, statistically significant, and genetic as a risk factor.

Even candidate genes are highly useful for studying known variants, there may still be unidentified genes that may adversely affect soft tissue injuries in different populations. As a result of the genetic tests carried out within the scope of the thesis, different polymorphisms were tested for 3 different gene regions, each using a separate sequencing method. These methods include qPCR analysis, PCR analysis & gel electrophoresis and Sanger sequencing. The reason for using 3 different methods varies in terms of the length of the region to be sequenced, whether the polymorphism includes deletion/insertion, and ease of use.

In the study, sequencing different intron variants close to the selected variant of rs1800012 with Sanger sequencing and combining them with data from athletes was

investigated to provide new variants associated with the phenotype in Turkish athletes. With the increase in case-control studies among these studies, it is important to investigate the effect of gene regions in different populations on the phenotype and to find new interactions between genes according to injury types. In the NCBI database, there were 22 previously associated Single nucleotide variants in the targeted 1000 bp region, which 3 of the variants had a 1000 Genome reference population score (Cartwright et al. 2015). In this thesis, the aim was to explain whether similar polymorphisms could be observed in the football cohort by performing multiple sequence alignments and analyzing the Sanger results one by one. However, there was no significant SNP found in the cohort which is previously associated with *COL1A1* gene.

For the sequencing of *ACE* and *ACTN3* genes for determining the muscle type variation among athletes, qPCR and PCR analysis was done to determine alleles and frequencies among football players. The physical data obtained were evaluated statistically, and our study proposal for soft tissue injuries will contribute to the prevention and rehabilitation of injuries. Population statistics showed that increased professional sports age and age has a higher risk of Hamstring injury. In addition to genetics, many other environmental factors (loading, severity, intensity, duration, place of training, temperature, etc.) are effective in being more prone to injuries in a sport, but the most important goal should be to reduce injury rates with precautions suitable for injury mechanisms. Also players encounter with injury when they play higher number of weekly matches than their teammates. It will contribute to the training of trainers in preparing their training programs to eliminate the deficiencies and to the training of athletes who respond better to the loads. None of the genetic variants were associated with Nordbord scores. Only *COL1A1* rs1800012 variant was marginally associated with overall soft tissue injury.

In the meta-analysis, *COL1A1* gene variant of rs1800012 in Caucasian cohort, G allele is associated with overall soft tissue injuries and has a positive effect in injury occurrence. To have T allele is protective against all studies in the cohort (Cartwright et al. 2015; Ryan-Moore et al. 2020; C. Wang et al. 2017).

It has also been tested in studies conducted abroad that athletes who have been genetically scored and matched with the right training (for example, strength genotype with high-intensity training or endurance genotype with low-intensity training) have also been

tested in studies conducted abroad, and the results observed up to 7 times increase in performance in athletes who turn to the appropriate training program.

To conclude, the aim was to test the hypothesis that athletes who have been genetically determined and matched accordingly with the correct training show greater improvement. With this setup, the upcoming studies will be one of the comprehensive studies in sports genetics studies conducted in Turkey, which includes scoring, enables quantitative measurements, and combines genetic factors with physical tests in the international literature. As a result of muscle type associations made with genetic analyzes, there are studies that direct athletes to the right training program and increase their performance.

REFERENCES

Ahmetov, II, E. S. Egorova, L. J. Gabdrakhmanova, and O. N. Fedotovskaya. 2016. "Genes and Athletic Performance: An Update." *Med Sport Sci* 61: 41-54. https://doi.org/10.1159/000445240. https://www.ncbi.nlm.nih.gov/pubmed/27287076.

Ahmetov, II, and O. N. Fedotovskaya. 2015. "Current Progress in Sports Genomics." *Adv Clin Chem* 70: 247-314. https://doi.org/10.1016/bs.acc.2015.03.003. https://www.ncbi.nlm.nih.gov/pubmed/26231489.

Ahmetov, II, and V. A. Rogozkin. 2009. "Genes, athlete status and training -- An overview." *Med Sport Sci* 54: 43-71. https://doi.org/10.1159/000235696. https://www.ncbi.nlm.nih.gov/pubmed/19696507.

Ahmetov, I., N. Kulemin, D. Popov, V. Naumov, E. Akimov, Y. Bravy, E. Egorova, A. Galeeva, E. Generozov, E. Kostryukova, A. Larin, Lj Mustafina, E. Ospanova, A. Pavlenko, L. Starnes, P. Zmijewski, D. Alexeev, O. Vinogradova, and V. Govorun. 2015. "Genome-wide association study identifies three novel genetic markers associated with elite endurance performance." *Biol Sport* 32 (1): 3-9. https://doi.org/10.5604/20831862.1124568. https://www.ncbi.nlm.nih.gov/pubmed/25729143.

Al-Khelaifi, F., I. Diboun, F. Donati, F. Botre, D. Abraham, A. Hingorani, O. Albagha, C. Georgakopoulos, K. Suhre, N. A. Yousri, and M. A. Elrayess. 2019. "Metabolic GWAS of elite athletes reveals novel genetically-influenced metabolites associated with athletic performance." *Sci Rep* 9 (1): 19889. https://doi.org/10.1038/s41598-019-56496-7. https://www.ncbi.nlm.nih.gov/pubmed/31882771.

Appel, M., K. Zentgraf, K. Kruger, and K. Alack. 2021. "Effects of Genetic Variation on Endurance Performance, Muscle Strength, and Injury Susceptibility in Sports: A Systematic Review." *Front Physiol* 12: 694411. https://doi.org/10.3389/fphys.2021.694411. https://www.ncbi.nlm.nih.gov/pubmed/34366884.

Baltazar-Martins, G., J. Gutierrez-Hellin, M. Aguilar-Navarro, C. Ruiz-Moreno, V. Moreno-Perez, A. Lopez-Samanes, R. Dominguez, and J. Del Coso. 2020. "Effect of ACTN3 Genotype on Sports Performance, Exercise-Induced Muscle Damage, and Injury Epidemiology." *Sports (Basel)* 8 (7). https://doi.org/10.3390/sports8070099. https://www.ncbi.nlm.nih.gov/pubmed/32668587.

Ben-Zaken, S., A. Eliakim, D. Nemet, M. Rabinovich, E. Kassem, and Y. Meckel. 2015. "ACTN3 Polymorphism: Comparison Between Elite Swimmers and Runners." *Sports Med Open* 1 (1): 13. https://doi.org/10.1186/s40798-015-0023-y. https://www.ncbi.nlm.nih.gov/pubmed/26284168.

Ben-Zaken, S., Y. Meckel, D. Nemet, E. Kassem, and A. Eliakim. 2021. "Genetic Basis for the Dominance of Israeli Long-Distance Runners of Ethiopian Origin." *J Strength Cond Res* 35 (7): 1885-1896. https://doi.org/10.1519/JSC.0000000000002989. https://www.ncbi.nlm.nih.gov/pubmed/30741858.

Brazier, J., M. Antrobus, G. K. Stebbings, S. H. Day, S. M. Heffernan, M. J. Cross, and A. G. Williams. 2019. "Tendon and Ligament Injuries in Elite Rugby: The Potential Genetic Influence." *Sports (Basel)* 7 (6). https://doi.org/10.3390/sports7060138. https://www.ncbi.nlm.nih.gov/pubmed/31167482.

Broos, S., L. Malisoux, D. Theisen, R. van Thienen, M. Ramaekers, C. Jamart, L. Deldicque, M. A. Thomis, and M. Francaux. 2016. "Evidence for ACTN3 as a Speed Gene in Isolated Human Muscle Fibers." *PLoS One* 11 (3): e0150594. https://doi.org/10.1371/journal.pone.0150594. https://www.ncbi.nlm.nih.gov/pubmed/26930663.

Buchheit, M., Y. Cholley, M. Nagel, and N. Poulos. 2016. "The Effect of Body Mass on Eccentric Knee-Flexor Strength Assessed With an Instrumented Nordic Hamstring Device (Nordbord) in Football Players." *Int J Sports Physiol Perform* 11 (6): 721-726. https://doi.org/10.1123/ijspp.2015-0513. https://www.ncbi.nlm.nih.gov/pubmed/26638728.

Cartwright, R., A. C. Kirby, K. A. Tikkinen, A. Mangera, G. Thiagamoorthy, P. Rajan, J. Pesonen, C. Ambrose, J. Gonzalez-Maffe, P. Bennett, T. Palmer, A. Walley, M. R. Jarvelin, C. Chapple, and V. Khullar. 2015. "Systematic review and metaanalysis of

genetic association studies of urinary symptoms and prolapse in women." *Am J Obstet Gynecol* 212 (2): 199 e1-24. https://doi.org/10.1016/j.ajog.2014.08.005. https://www.ncbi.nlm.nih.gov/pubmed/25111588.

Cieszczyk, P., K. Krupecki, A. Maciejewska, and M. Sawczuk. 2009. "The angiotensin converting enzyme gene I/D polymorphism in Polish rowers." *Int J Sports Med* 30 (8): 624-7. https://doi.org/10.1055/s-0029-1202825. https://www.ncbi.nlm.nih.gov/pubmed/19455482.

Clos, E., R. Pruna, M. Lundblad, R. Artells, and J. Esquirol Caussa. 2019. "ACTN3 single nucleotide polymorphism is associated with non-contact musculoskeletal soft-tissue injury incidence in elite professional football players." *Knee Surg Sports Traumatol Arthrosc* 27 (12): 4055-4061. https://doi.org/10.1007/s00167-019-05381-x. https://www.ncbi.nlm.nih.gov/pubmed/30721342.

Collins, M. 2010. "Genetic risk factors for soft-tissue injuries 101: a practical summary to help clinicians understand the role of genetics and 'personalised medicine'." *Br J Sports Med* 44 (13): 915-7. https://doi.org/10.1136/bjsm.2009.058040. https://www.ncbi.nlm.nih.gov/pubmed/19553227.

Collins, M., A. V. September, and M. Posthumus. 2015. "Biological variation in musculoskeletal injuries: current knowledge, future research and practical implications." *Br J Sports Med* 49 (23): 1497-503. https://doi.org/10.1136/bjsports-2015-095180. https://www.ncbi.nlm.nih.gov/pubmed/26504180.

de la Iglesia, R., I. Espinosa-Salinas, F. J. Lopez-Silvarrey, J. J. Ramos-Alvarez, J. C. Segovia, G. Colmenarejo, E. Borregon-Rivilla, H. Marcos-Pasero, E. Aguilar-Aguilar, V. Loria-Kohen, G. Reglero, and A. Ramirez-de Molina. 2020. "A Potential Endurance Algorithm Prediction in the Field of Sports Performance." *Front Genet* 11: 711. https://doi.org/10.3389/fgene.2020.00711. https://www.ncbi.nlm.nih.gov/pubmed/32849773.

Del Coso, J., D. Hiam, P. Houweling, L. M. Perez, N. Eynon, and A. Lucia. 2019. "More than a 'speed gene': ACTN3 R577X genotype, trainability, muscle damage, and the risk for injuries." *Eur J Appl Physiol* 119 (1): 49-60. https://doi.org/10.1007/s00421-018-4010-0. https://www.ncbi.nlm.nih.gov/pubmed/30327870.

Del Coso, J., V. Moreno, J. Gutierrez-Hellin, G. Baltazar-Martins, C. Ruiz-Moreno, M. Aguilar-Navarro, B. Lara, and A. Lucia. 2019. "ACTN3 R577X Genotype and Exercise Phenotypes in Recreational Marathon Runners." *Genes (Basel)* 10 (6). https://doi.org/10.3390/genes10060413. https://www.ncbi.nlm.nih.gov/pubmed/31146466.

Del Coso, J., J. J. Salinero, B. Lara, C. Gallo-Salazar, F. Areces, C. Puente, and D. Herrero. 2017. "ACTN3 X-allele carriers had greater levels of muscle damage during a half-ironman." *Eur J Appl Physiol* 117 (1): 151-158. https://doi.org/10.1007/s00421-016-3507-7. https://www.ncbi.nlm.nih.gov/pubmed/27913923.

Ehlert, T., P. Simon, and D. A. Moser. 2013. "Epigenetics in sports." *Sports Med* 43 (2): 93-110. https://doi.org/10.1007/s40279-012-0012-y. https://www.ncbi.nlm.nih.gov/pubmed/23329609.

Eken, B , Sercan, C , Kaya, D , Ulucan, K 2018. "Darbeye Bağlı Olmayan Yumuşak Doku Yaralanmalarında Moleküler Yaklaşım." *Eurasian Research in Sport Science* 3: 15-25.

Engebretsen, A. H., G. Myklebust, I. Holme, L. Engebretsen, and R. Bahr. 2008. "Prevention of injuries among male soccer players: a prospective, randomized intervention study targeting players with previous injuries or reduced function." *Am J Sports Med* 36 (6): 1052-60. https://doi.org/10.1177/0363546508314432.

Eroglu, O., R. Zileli, M. A. Nalbant, and K. Ulucan. 2018. "Prevalence of alpha actinin-3 gene (ACTN3) R577X and angiotensin converting enzyme (ACE) insertion / deletion gene polymorphisms in national and amateur Turkish athletes." *Cell Mol Biol (Noisy-le-grand)* 64 (5): 24-28. https://www.ncbi.nlm.nih.gov/pubmed/29729690.

Eroğlu, O., Zileli, R. . (2015). " Genetik Faktörlerin Sportif Performansa Etkisi." *Uluslararası Spor, Egzersiz ve Antrenman Bilimi Dergisi* 1: 63-76.

Eynon, N., E. D. Hanson, A. Lucia, P. J. Houweling, F. Garton, K. N. North, and D. J. Bishop. 2013. "Genes for elite power and sprint performance: ACTN3 leads the way." *Sports Med* 43 (9): 803-17. https://doi.org/10.1007/s40279-013-0059-4. https://www.ncbi.nlm.nih.gov/pubmed/23681449.

Fawcett, K. A., and I. Barroso. 2010. "The genetics of obesity: FTO leads the way." *Trends Genet* 26 (6): 266-74. https://doi.org/10.1016/j.tig.2010.02.006. https://www.ncbi.nlm.nih.gov/pubmed/20381893.

Fedotovskaya, O. N., L. J. Mustafina, D. V. Popov, O. L. Vinogradova, and Ahmetov, II. 2014. "A common polymorphism of the MCT1 gene and athletic performance." *Int J Sports Physiol Perform* 9 (1): 173-80. https://doi.org/10.1123/ijspp.2013-0026. https://www.ncbi.nlm.nih.gov/pubmed/23628675.

Feeley, B. T., S. Kennelly, R. P. Barnes, M. S. Muller, B. T. Kelly, S. A. Rodeo, and R. F. Warren. 2008. "Epidemiology of National Football League training camp injuries from 1998 to 2007." *Am J Sports Med* 36 (8): 1597-603. https://doi.org/10.1177/0363546508316021. https://www.ncbi.nlm.nih.gov/pubmed/18443276.

Feito, Y., K. M. Heinrich, S. J. Butcher, and W. S. C. Poston. 2018. "High-Intensity Functional Training (HIFT): Definition and Research Implications for Improved Fitness." *Sports (Basel)* 6 (3). https://doi.org/10.3390/sports6030076.

Fleming, I. 2006. "Signaling by the angiotensin-converting enzyme." *Circ Res* 98 (7): 887-96. https://doi.org/10.1161/01.RES.0000217340.40936.53. https://www.ncbi.nlm.nih.gov/pubmed/16614314.

Frazer, K. A., S. S. Murray, N. J. Schork, and E. J. Topol. 2009. "Human genetic variation and its contribution to complex traits." *Nat Rev Genet* 10 (4): 241-51. https://doi.org/10.1038/nrg2554. https://www.ncbi.nlm.nih.gov/pubmed/19293820.

Gabbett, T. J., S. Ullah, and C. F. Finch. 2012. "Identifying risk factors for contact injury in professional rugby league players--application of a frailty model for recurrent injury." *J Sci Med Sport* 15 (6): 496-504. https://doi.org/10.1016/j.jsams.2012.03.017. https://www.ncbi.nlm.nih.gov/pubmed/22748762.

Gibbon, A., H. Hobbs, W. van der Merwe, S. M. Raleigh, J. Cook, C. J. Handley, M. Posthumus, M. Collins, and A. V. September. 2017. "The MMP3 gene in musculoskeletal soft tissue injury risk profiling: A study in two independent sample groups." *J Sports Sci* 35 (7): 655-662.

https://doi.org/10.1080/02640414.2016.1183806. https://www.ncbi.nlm.nih.gov/pubmed/27211292.

Gibson, W. T. 2009. "Genetic association studies for complex traits: relevance for the sports medicine practitioner." *Br J Sports Med* 43 (5): 314-6. https://doi.org/10.1136/bjsm.2008.052191. https://www.ncbi.nlm.nih.gov/pubmed/19066182.

Gineviciene, V., A. Jakaitiene, M. O. Aksenov, A. V. Aksenova, A. M. Druzhevskaya, I. V. Astratenkova, E. S. Egorova, L. J. Gabdrakhmanova, L. Tubelis, V. Kucinskas, and A. Utkus. 2016. "Association analysis of ACE, ACTN3 and PPARGC1A gene polymorphisms in two cohorts of European strength and power athletes." *Biol Sport* 33 (3): 199-206. https://doi.org/10.5604/20831862.1201051. https://www.ncbi.nlm.nih.gov/pubmed/27601773.

Gineviciene, V., A. Pranculis, A. Jakaitiene, K. Milasius, and V. Kucinskas. 2011. "Genetic variation of the human ACE and ACTN3 genes and their association with functional muscle properties in Lithuanian elite athletes." *Medicina (Kaunas)* 47 (5): 284-90. https://www.ncbi.nlm.nih.gov/pubmed/21956137.

Goode, M. R., S. Y. Cheong, N. Li, W. C. Ray, and C. W. Bartlett. 2014. "Collection and extraction of saliva DNA for next generation sequencing." *J Vis Exp* (90). https://doi.org/10.3791/51697. https://www.ncbi.nlm.nih.gov/pubmed/25225892.

Griffin, L. Y., M. J. Albohm, E. A. Arendt, R. Bahr, B. D. Beynnon, M. Demaio, R. W. Dick, L. Engebretsen, W. E. Garrett, Jr., J. A. Hannafin, T. E. Hewett, L. J. Huston, M. L. Ireland, R. J. Johnson, S. Lephart, B. R. Mandelbaum, B. J. Mann, P. H. Marks, S. W. Marshall, G. Myklebust, F. R. Noyes, C. Powers, C. Shields, Jr., S. J. Shultz, H. Silvers, J. Slauterbeck, D. C. Taylor, C. C. Teitz, E. M. Wojtys, and B. Yu. 2006. "Understanding and preventing noncontact anterior cruciate ligament injuries: a review of the Hunt Valley II meeting, January 2005." *Am J Sports Med* 34 (9): 1512-32. https://doi.org/10.1177/0363546506286866. https://www.ncbi.nlm.nih.gov/pubmed/16905673.

Guilherme, Jplf, R. Bertuzzi, A. E. Lima-Silva, A. D. C. Pereira, and A. H. Lancha Junior. 2018. "Analysis of sports-relevant polymorphisms in a large Brazilian cohort of top-

level athletes." *Ann Hum Genet* 82 (5): 254-264. https://doi.org/10.1111/ahg.12248. https://www.ncbi.nlm.nih.gov/pubmed/29603120.

Guo, R., Z. Ji, S. Gao, A. Aizezi, Y. Fan, Z. Wang, and K. Ning. 2022. "Association of COL5A1 gene polymorphisms and musculoskeletal soft tissue injuries: a meta-analysis based on 21 observational studies." *J Orthop Surg Res* 17 (1): 129. https://doi.org/10.1186/s13018-022-03020-9. https://www.ncbi.nlm.nih.gov/pubmed/35241120.

Guth, L. M., and S. M. Roth. 2013. "Genetic influence on athletic performance." *Curr Opin Pediatr* 25 (6): 653-8. https://doi.org/10.1097/MOP.0b013e3283659087. https://www.ncbi.nlm.nih.gov/pubmed/24240283.

Henderson, G., C. A. Barnes, and M. D. Portas. 2010. "Factors associated with increased propensity for hamstring injury in English Premier League soccer players." *J Sci Med Sport* 13 (4): 397-402. https://doi.org/10.1016/j.jsams.2009.08.003. https://www.ncbi.nlm.nih.gov/pubmed/19800844.

Hogberg, J., E. Bergentoft, R. Piussi, M. Wernbom, S. Beischer, R. Simonson, C. Senorski, R. Thomee, and E. Hamrin Senorski. 2022. "Persistent knee flexor strength deficits identified through the NordBord eccentric test not seen with "gold standard" isokinetic concentric testing during the first year after anterior cruciate ligament reconstruction with a hamstring tendon autograft." *Phys Ther Sport* 55: 119-124. https://doi.org/10.1016/j.ptsp.2022.03.004. https://www.ncbi.nlm.nih.gov/pubmed/35339867.

Houweling, P. J., Y. D. Berman, N. Turner, K. G. R. Quinlan, J. T. Seto, N. Yang, M. Lek, D. G. Macarthur, G. Cooney, and K. N. North. 2017. "Exploring the relationship between alpha-actinin-3 deficiency and obesity in mice and humans." *Int J Obes (Lond)* 41 (7): 1154-1157. https://doi.org/10.1038/ijo.2017.72. https://www.ncbi.nlm.nih.gov/pubmed/28293018.

Jones, N., J. Kiely, B. Suraci, D. J. Collins, D. de Lorenzo, C. Pickering, and K. A. Grimaldi. 2016. "A genetic-based algorithm for personalized resistance training." *Biol Sport* 33 (2): 117-26. https://doi.org/10.5604/20831862.1198210. https://www.ncbi.nlm.nih.gov/pubmed/27274104.

---. 2017. "A response to letter to the editor: A genetic-based algorithm for personalized resistance training." *Biol Sport* 34 (1): 35-37. https://doi.org/10.5114/biolsport.2017.63386. https://www.ncbi.nlm.nih.gov/pubmed/28416895.

Kang, H. J., C. H. Kim, D. S. Park, S. Y. Choi, D. H. Lee, H. S. Nam, J. G. Hur, and J. H. Woo. 2012. "The Impacts of ACE Activity according to ACE I/D Polymorphisms on Muscular Functions of People Aged 65." *Ann Rehabil Med* 36 (4): 433-46. https://doi.org/10.5535/arm.2012.36.4.433. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3438409/pdf/arm-36-433.pdf.

Kasapis, C., and P. D. Thompson. 2005. "The effects of physical activity on serum C-reactive protein and inflammatory markers: a systematic review." *J Am Coll Cardiol* 45 (10): 1563-9. https://doi.org/10.1016/j.jacc.2004.12.077. https://www.ncbi.nlm.nih.gov/pubmed/15893167.

Kim, S. K., T. R. Roos, A. K. Roos, J. P. Kleimeyer, M. A. Ahmed, G. T. Goodlin, M. Fredericson, J. P. Ioannidis, A. L. Avins, and J. L. Dragoo. 2017. "Genome-wide association screens for Achilles tendon and ACL tears and tendinopathy." *PLoS One* 12 (3): e0170422. https://doi.org/10.1371/journal.pone.0170422. https://www.ncbi.nlm.nih.gov/pubmed/28358823.

KOKU, F. E. . 2015. "Sportif Performansin Genetik İle İlişkisi." *Turkish Journal of Sports Medicine* volume 50 (1).

Kujala, U. M., T. Palviainen, P. Pesonen, K. Waller, E. Sillanpaa, M. Niemela, M. Kangas, H. Vaha-Ypya, H. Sievanen, R. Korpelainen, T. Jamsa, M. Mannikko, and J. Kaprio. 2020. "Polygenic Risk Scores and Physical Activity." *Med Sci Sports Exerc* 52 (7): 1518-1524. https://doi.org/10.1249/MSS.0000000000002290. https://www.ncbi.nlm.nih.gov/pubmed/32049886.

Lambert, G., D. Tsinajinnie, and D. Duggan. 2013. "Single nucleotide polymorphism genotyping using BeadChip microarrays." *Curr Protoc Hum Genet* Chapter 2: Unit 2.9. https://doi.org/10.1002/0471142905.hg0209s78.

Larruskain, J., D. Celorrio, I. Barrio, A. Odriozola, S. M. Gil, J. R. Fernandez-Lopez, R. Nozal, I. Ortuzar, J. A. Lekue, and J. M. Aznar. 2018. "Genetic Variants and Hamstring Injury in Soccer: An Association and Validation Study." *Med Sci Sports*

Exerc 50 (2): 361-368. https://doi.org/10.1249/MSS.0000000000001434. https://www.ncbi.nlm.nih.gov/pubmed/28976491.

Leonska-Duniec, A., A. Grzywacz, Z. Jastrzebski, A. Jazdzewska, E. Lulinska-Kuklik, W. Moska, K. Leznicka, K. Ficek, A. Rzeszutko, M. Dornowski, and P. Cieszczyk. 2018. "ADIPOQ polymorphisms are associated with changes in obesity-related traits in response to aerobic training programme in women." *Biol Sport* 35 (2): 165-173. https://doi.org/10.5114/biolsport.2018.72762. https://www.ncbi.nlm.nih.gov/pubmed/30455545.

Lopez-Leon, S., C. Tuvblad, and D. A. Forero. 2016. "Sports genetics: the PPARA gene and athletes' high ability in endurance sports. A systematic review and meta-analysis." *Biol Sport* 33 (1): 3-6. https://doi.org/10.5604/20831862.1180170. https://www.ncbi.nlm.nih.gov/pubmed/26985127.

Lulinska-Kuklik, E., M. N. Laguette, W. Moska, M. Weber-Rajek, K. Ficek, R. Puchala, P. Cieszczyk, M. Sawczuk, A. V. September, and A. Maciejewska-Skrendo. 2019. "Are TNC gene variants associated with anterior cruciate ligament rupture susceptibility?" *J Sci Med Sport* 22 (4): 408-412. https://doi.org/10.1016/j.jsams.2018.10.003. https://www.ncbi.nlm.nih.gov/pubmed/30528246.

Lulinska-Kuklik, E., M. Rahim, W. Moska, E. Maculewicz, M. Kaczmarczyk, A. Maciejewska-Skrendo, K. Ficek, P. Cieszczyk, A. V. September, and M. Sawczuk. 2019. "Are MMP3, MMP8 and TIMP2 gene variants associated with anterior cruciate ligament rupture susceptibility?" *J Sci Med Sport* 22 (7): 753-757. https://doi.org/10.1016/j.jsams.2019.01.014. https://www.ncbi.nlm.nih.gov/pubmed/30755371.

Maffulli, N., K. Margiotti, U. G. Longo, M. Loppini, V. M. Fazio, and V. Denaro. 2013. "The genetics of sports injuries and athletic performance." *Muscles Ligaments Tendons J* 3 (3): 173-89. https://www.ncbi.nlm.nih.gov/pubmed/24367777.

Neto, Slda, J. J. B. Herrera, T. S. Rosa, S. S. de Almeida, G. C. B. Silva, C. E. S. Ferreira, M. A. P. Dos Santos, V. O. Silvino, and G. F. de Melo. 2020. "Interaction Between ACTN3 (R577X), ACE (I/D), and BDKRB2 (-9/+9) Polymorphisms and Endurance Phenotypes in Brazilian Long-Distance Swimmers." *J Strength Cond Res*.

https://doi.org/10.1519/JSC.0000000000003685. https://www.ncbi.nlm.nih.gov/pubmed/32639377.

Ntanasis-Stathopoulos, J., J. G. Tzanninis, A. Philippou, and M. Koutsilieris. 2013. "Epigenetic regulation on gene expression induced by physical exercise." *J Musculoskelet Neuronal Interact* 13 (2): 133-46. https://www.ncbi.nlm.nih.gov/pubmed/23728100.

O'Connell, K., H. Knight, K. Ficek, A. Leonska-Duniec, A. Maciejewska-Karlowska, M. Sawczuk, M. Stepien-Slodkowska, D. O'Cuinneagain, W. van der Merwe, M. Posthumus, P. Cieszczyk, and M. Collins. 2015. "Interactions between collagen gene variants and risk of anterior cruciate ligament rupture." *Eur J Sport Sci* 15 (4): 341-50. https://doi.org/10.1080/17461391.2014.936324. https://www.ncbi.nlm.nih.gov/pubmed/25073002.

Opar, D. A., M. D. Williams, and A. J. Shield. 2012. "Hamstring strain injuries: factors that lead to injury and re-injury." *Sports Med* 42 (3): 209-26. https://doi.org/10.2165/11594800-000000000-00000.

Opitz, D., E. Lenzen, A. Opiolka, M. Redmann, M. Hellmich, W. Bloch, K. Brixius, and C. Brinkmann. 2015. "Endurance training alters basal erythrocyte MCT-1 contents and affects the lactate distribution between plasma and red blood cells in T2DM men following maximal exercise." *Can J Physiol Pharmacol* 93 (6): 413-9. https://doi.org/10.1139/cjpp-2014-0467. https://www.ncbi.nlm.nih.gov/pubmed/25844530.

Pareja-Galeano, H., F. Sanchis-Gomar, and J. L. Garcia-Gimenez. 2014. "Physical exercise and epigenetic modulation: elucidating intricate mechanisms." *Sports Med* 44 (4): 429-36. https://doi.org/10.1007/s40279-013-0138-6. https://www.ncbi.nlm.nih.gov/pubmed/24399634.

Patil, N., N. Nouri, L. McAllister, H. Matsukaki, and T. Ryder. 2001. "Single-nucleotide polymorphism genotyping using microarrays." *Curr Protoc Hum Genet* Chapter 2: Unit 2.9. https://doi.org/10.1002/0471142905.hg0209s27.

Pearson, H. 2006. "Genetics: what is a gene?" *Nature* 441 (7092): 398-401. https://doi.org/10.1038/441398a. https://www.ncbi.nlm.nih.gov/pubmed/16724031.

Pehlivan, Z. 2013. "OKULLARARASI SPOR YARIŞMALARINA KATILANLARIN SPORU BIRAKMA VE DEVAM ETME NEDENLERİ." *Spor Bilimleri Dergisi* 24 (3): 209-225.

Pickering, C., and J. Kiely. 2017. "ACTN3: More than Just a Gene for Speed." *Front Physiol* 8: 1080. https://doi.org/10.3389/fphys.2017.01080. https://www.ncbi.nlm.nih.gov/pubmed/29326606.

Pickering, C., J. Kiely, B. Suraci, and D. Collins. 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. https://www.ncbi.nlm.nih.gov/pubmed/30485313.

Pickering, C., B. Suraci, E. A. Semenova, E. A. Boulygina, E. S. Kostryukova, N. A. Kulemin, O. V. Borisov, S. A. Khabibova, A. K. Larin, A. V. Pavlenko, E. V. Lyubaeva, D. V. Popov, E. A. Lysenko, T. F. Vepkhvadze, E. M. Lednev, A. Leonska-Duniec, B. Pajak, J. Chycki, W. Moska, E. Lulinska-Kuklik, M. Dornowski, A. Maszczyk, B. Bradley, A. Kana-Ah, P. Cieszczyk, E. V. Generozov, and Ahmetov, II. 2019. "A Genome-Wide Association Study of Sprint Performance in Elite Youth Football Players." *J Strength Cond Res* 33 (9): 2344-2351. https://doi.org/10.1519/JSC.0000000000003259. https://www.ncbi.nlm.nih.gov/pubmed/31343553.

Pimenta, I., H. Mateus, S. Rodrigues-Manica, R. Pinheiro-Torres, A. Neto, L. Domingues, C. Lage Crespo, A. Sardoo, P. Machado, J. C. Branco, S. N. Silva, and F. M. Pimentel-Santos. 2021. "The Effect of ACTN3 and VDR Polymorphisms on Skeletal Muscle Performance in Axial Spondyloarthropathies." *Front Genet* 12: 688984. https://doi.org/10.3389/fgene.2021.688984. https://www.ncbi.nlm.nih.gov/pubmed/34456969.

Pranckeviciene, E., V. Gineviciene, A. Jakaitiene, L. Januska, and A. Utkus. 2021. "Total Genotype Score Modelling of Polygenic Endurance-Power Profiles in Lithuanian Elite Athletes." *Genes (Basel)* 12 (7). https://doi.org/10.3390/genes12071067. https://www.ncbi.nlm.nih.gov/pubmed/34356082.

Pruna, R., R. Artells, J. Ribas, B. Montoro, F. Cos, C. Munoz, G. Rodas, and N. Maffulli. 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 Musculoskelet Disord* 14: 221. https://doi.org/10.1186/1471-2474-14-221. https://www.ncbi.nlm.nih.gov/pubmed/23890452.

Puthucheary, Z., J. R. Skipworth, J. Rawal, M. Loosemore, K. Van Someren, and H. E. Montgomery. 2011. "The ACE gene and human performance: 12 years on." *Sports Med* 41 (6): 433-48. https://doi.org/10.2165/11588720-000000000-00000. https://www.ncbi.nlm.nih.gov/pubmed/21615186.

Reihmane, D., and F. Dela. 2014. "Interleukin-6: possible biological roles during exercise." *Eur J Sport Sci* 14 (3): 242-50. https://doi.org/10.1080/17461391.2013.776640. https://www.ncbi.nlm.nih.gov/pubmed/24655147.

Ryan-Moore, E., Y. Mavrommatis, and M. Waldron. 2020. "Systematic Review and Meta-Analysis of Candidate Gene Association Studies With Fracture Risk in Physically Active Participants." *Front Genet* 11: 551. https://doi.org/10.3389/fgene.2020.00551.

Siegrist, M. 2008. "[Role of physical activity in the prevention of osteoporosis]." *Med Monatsschr Pharm* 31 (7): 259-64. https://www.ncbi.nlm.nih.gov/pubmed/18808074.

Stastny, P., M. Lehnert, M. De Ste Croix, M. Petr, Z. Svoboda, E. Maixnerova, R. Varekova, M. Botek, M. Petrek, L. Kocourkova, and P. Cieszczyk. 2019. "Effect of COL5A1, GDF5, and PPARA Genes on a Movement Screen and Neuromuscular Performance in Adolescent Team Sport Athletes." *J Strength Cond Res* 33 (8): 2057-2065. https://doi.org/10.1519/JSC.0000000000003142. https://www.ncbi.nlm.nih.gov/pubmed/30985523.

Szulkin, R., T. Whitington, M. Eklund, M. Aly, R. A. Eeles, D. Easton, Z. S. Kote-Jarai, A. Amin Al Olama, S. Benlloch, K. Muir, G. G. Giles, M. C. Southey, L. M. Fitzgerald, B. E. Henderson, F. Schumacher, C. A. Haiman, J. Schleutker, T. Wahlfors, T. L. Tammela, B. G. Nordestgaard, T. J. Key, R. C. Travis, D. E. Neal, J. L. Donovan, F. C. Hamdy, P. Pharoah, N. Pashayan, K. T. Khaw, J. L. Stanford, S. N. Thibodeau, S. K. McDonnell, D. J. Schaid, C. Maier, W. Vogel, M. Luedeke,

K. Herkommer, A. S. Kibel, C. Cybulski, J. Lubinski, W. Kluzniak, L. Cannon-Albright, H. Brenner, K. Butterbach, C. Stegmaier, J. Y. Park, T. Sellers, H. Y. Lin, C. Slavov, R. Kaneva, V. Mitev, J. Batra, J. A. Clements, BioResource Australian Prostate Cancer, A. Spurdle, M. R. Teixeira, P. Paulo, S. Maia, H. Pandha, A. Michael, A. Kierzek, Consortium Practical, H. Gronberg, and F. Wiklund. 2015. "Prediction of individual genetic risk to prostate cancer using a polygenic score." *Prostate* 75 (13): 1467-74. https://doi.org/10.1002/pros.23037. https://www.ncbi.nlm.nih.gov/pubmed/26177737.

Tharabenjasin, P., N. Pabalan, and H. Jarjanazi. 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. https://www.ncbi.nlm.nih.gov/pubmed/31145768.

Ulucan, K., C. Sercan, and T. Biyikli. 2015. "Distribution of Angiotensin-1 Converting Enzyme Insertion/Deletion and alpha-Actinin-3 Codon 577 Polymorphisms in Turkish Male Soccer Players." *Genet Epigenet* 7: 1-4. https://doi.org/10.4137/GEG.S31479. https://www.ncbi.nlm.nih.gov/pubmed/26448692.

Valdivieso, P., D. Vaughan, E. Laczko, M. Brogioli, S. Waldron, J. Rittweger, and M. Fluck. 2017. "The Metabolic Response of Skeletal Muscle to Endurance Exercise Is Modified by the ACE-I/D Gene Polymorphism and Training State." *Front Physiol* 8: 993. https://doi.org/10.3389/fphys.2017.00993. https://www.ncbi.nlm.nih.gov/pubmed/29311951.

Vina, J., F. Sanchis-Gomar, V. Martinez-Bello, and M. C. Gomez-Cabrera. 2012. "Exercise acts as a drug; the pharmacological benefits of exercise." *Br J Pharmacol* 167 (1): 1-12. https://doi.org/10.1111/j.1476-5381.2012.01970.x. https://www.ncbi.nlm.nih.gov/pubmed/22486393.

Vincent, B., K. De Bock, M. Ramaekers, E. Van den Eede, M. Van Leemputte, P. Hespel, and M. A. Thomis. 2007. "ACTN3 (R577X) genotype is associated with fiber type distribution." *Physiol Genomics* 32 (1): 58-63. https://doi.org/10.1152/physiolgenomics.00173.2007.

Wang, C., H. Li, K. Chen, B. Wu, and H. Liu. 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. https://www.ncbi.nlm.nih.gov/pubmed/28206959.

Wang, G., S. Padmanabhan, B. Wolfarth, N. Fuku, A. Lucia, Ahmetov, II, P. Cieszczyk, M. Collins, N. Eynon, V. Klissouras, A. Williams, and Y. Pitsiladis. 2013. "Genomics of elite sporting performance: what little we know and necessary advances." *Adv Genet* 84: 123-49. https://doi.org/10.1016/B978-0-12-407703-4.00004-9. https://www.ncbi.nlm.nih.gov/pubmed/24262098.

Wang, G., M. Tanaka, N. Eynon, K. N. North, A. G. Williams, M. Collins, C. N. Moran, S. L. Britton, N. Fuku, E. A. Ashley, V. Klissouras, A. Lucia, Ahmetov, II, E. de Geus, M. Alsayrafi, and Y. P. Pitsiladis. 2016. "The Future of Genomic Research in Athletic Performance and Adaptation to Training." *Med Sport Sci* 61: 55-67. https://doi.org/10.1159/000445241. https://www.ncbi.nlm.nih.gov/pubmed/27287077.

Williams, A. G., M. P. Rayson, M. Jubb, M. World, D. R. Woods, M. Hayward, J. Martin, S. E. Humphries, and H. E. Montgomery. 2000. "The ACE gene and muscle performance." *Nature* 403 (6770): 614. https://doi.org/10.1038/35001141. https://www.ncbi.nlm.nih.gov/pubmed/10688186.

Wilson, G. C., Y. Mavros, L. Tajouri, and M. F. Singh. 2019. "The Role of Genetic Profile in Functional Performance Adaptations to Exercise Training or Physical Activity: A Systematic Review of the Literature." *J Aging Phys Act* 27 (4): 594-616. https://doi.org/10.1123/japa.2018-0155. https://www.ncbi.nlm.nih.gov/pubmed/30676214.

Zhong, B., D. Huang, K. Ma, X. Deng, D. Shi, F. Wu, and Z. Shao. 2017. "Association of COL1A1 rs1800012 polymorphism with musculoskeletal degenerative diseases: a meta-analysis." *Oncotarget* 8 (43): 75488-75499. https://doi.org/10.18632/oncotarget.20797. https://www.ncbi.nlm.nih.gov/pubmed/29088884.

APPENDIX A

SEQUENCE INFORMATION

Reference sequence for *COL1A1* gene selected for Sanger sequencing;

GTAAGTCCCAAACTTTTGGGAGTGCAAGGATACTCTATATCGCGCCTTGCGCTT
GGTCCCGGGGGCCGCGGCTTAAAACGAGACGTGGATGATCCGGAGACTCGGGA
ATGGAAGGGAGATGATGAGGGCTCTTCCTCGGCGCCCTGAGACAGGAGGGAGC
TCACCCTGGGGCGAGGTTGGGGTTGAACGCGCCCCGGGAGCGGGAGGTGAGG
GTGGAGCGCGGCGTGAGTTGGTGCAAGAGAGAATCCCGAGCGCGCAACCGGG
GAAGTGGGGATCTGGGTGCAGAGTGAGGAAAGCACGTCGAAGATGGGATGGG
GGCGCCGAGCGGGGCATTTGAAGCCCAAGATGTAGAAGCAATCAGGAAGGCC
GTGGGATGATTCATAAGGAAAGATTGCCCTCTCTGCGGGCTAGAGTGTTGCTG
GGGCCGTGGGGGTGCTGGGCAGCCGCGGAGGGGGTGCGGAGCGTGGGCGGGT
GGAGGATGAGAAACTTTGGCGCGGACTCGGCGGGGCGGGGTCCTTGCGCCCCC
TGCTGACCGATGCTGAGCACTGCGTCTCCCGGTCCAACGCTTACTGGGGCAGG
AGCCGGAGCGGGAAGACCCGGGTTATTGCTGGGTGCGGACCCCCACCTCTAGA
TCTGGAAAGTAAAGCCAGGGATGGGGCAGCCCAAGCCTCTTAAAGAGGTAGTC
GGGCCGGTGAGGTCGGCCCCGCCCCGGCCCCATTGCTTAGCGTTGCCCGACAC
CTAGTGGCCGTCTGGGGAGCCGCTAGCGCGGTGGGAGTGGTTAGCTAACTTCT
GGACTATTTGCGGACTTTTTGGTTCTTTGGCTAAAAGTGACCTGGAGGCATTGG
CTGGCTTTGGGGGACTGGGGATGGCCCCGAGAGCGGGCTTTTAAGATGTCTAG
GTGCTGGAGGTTAGGGTGTCTCCTAATTTTGAGGTACATTTCAAGTCTTGGGGG
GGCCTCCCTTCCAATCAGCCGCTCCCATTCTCCTAGCCCCGCCCCCGCCACCCC
ACCTGCCCAGGGAATGGGGGCGGGATGAGGGCTGGACCTCCCTTCTCTCCTCC
CTCGCCCTCCTCCTGTCTCTACCACGCAAGCCACTCCCCACGAGCCTGCCCTCC
CGATGGGGCCCCTCCTATTCTCCCCCCGCCCTCCCCCTCTCACCCTGTGGTTTTT
ATTTCACTTGGCTTCAGCGCCAATGGGCTGAGGTTGGAGTTGGAAGCCACCGC
GGACTAAAGCTTTGTTTAAATTCCTGAGAACTGGAAAGAGTTACAGCCTCCCTG

GCCAGGCGCCTCGGCGCTGTCACCCGCGCTGATGAGGAGCAGGCGAGCTTTTA
AGGATTTGAGGAAAGAAGAACGGGGGGAGGGGCGGGAAGTGAAAAATCCAAG
TGTGCCTCTTAGACCCGGGGGAAAGGTGGTTAAGCTGGGGGTTGCAGTCACTA
CTGACAACGCCCCTCTTCCGCCTGTCCCAG