Thieme

Two Genetic Variants Associated with Plantar Fascial Disorders

Authors
Stuart K. Kim[1], John P.A. Ioannidis[2], Marwa A. Ahmed[3], Andrew L. Avins[4], John P. Kleimeyer[5], Michael Fredericson[5], Jason L. Dragoo[5]

Affiliations
1 Dept. Developmental Biology, Stanford University Medical Center, Stanford CA, USA
2 Dept. of Medicine, Stanford Prevention Research Center and Dept. of Health Research and Policy, Division of Epidemiology, Stanford University School of Medicine, and Dept. of Statistics, Stanford University School of Humanities and Sciences, Stanford CA, USA
3 Dept. Physical Medicine and Rehabilitation, Harvard Medical School, Boston MA, USA
4 Kaiser Permanente Northern California, Division of Research, Oakland, California, USA
5 Dept. Orthopaedic Surgery, Stanford University Medical Center, Stanford CA, USA

Key words
Genome-wide association study, plantar fasciitus, sports injury, plantar fibromatosis

Bibliography
DOI https://doi.org/10.1055/s-0044-100280
Published online: 13.3.2018
Int J Sports Med 2018; 39: 314–321

ISSN 0172-4622

Correspondence
Kim, Stuart
Dept. of Dev. Bio.
Stanford University Medical Center
Stanford, California 94305
United States
Tel.: +1/650/867 3189, Fax: +1/650/725 7739
stuartkm@stanford.edu

ABSTRACT

Plantar fascial disorder is comprised of plantar fasciitis and plantar fibromatosis. Plantar fasciitis is the most common cause of heel pain, especially for athletes involved in running and jumping sports. Plantar fibromatosis is a rare fibrous hyperproliferation of the deep connective tissue of the foot. To identify genetic loci associated with plantar fascial disorders, a genome-wide association screen was performed using publically available data from the Research Program in Genes, Environment and Health including 21,624 cases of plantar fascial disorders and 80,879 controls. One indel (chr5:118704153:D) and one SNP (rs62051384) showed an association with plantar fascial disorders at genome-wide significance ($p < 5 \times 10^{-8}$) with small effects (odds ratios = 0.93 and 1.07 per allele, respectively). The indel chr5:118704153:D is located within TNFAIP8 (encodes a protein induced by TNF alpha) and rs62051384 is located within WWP2 (which is involved in proteasomal degradation). These DNA variants may be informative in explaining why some individuals are at higher risk for plantar fascial disorders than others.

Introduction

Plantar fasciitis and plantar fibromatosis are two types of heel pain jointly referred to as plantar fascial disorders. Plantar fasciitis is the most common cause of acquired sub-calcaneal heel pain in adults with a prevalence of up to 20 % in the general population [1, 11]. Resulting from overuse and repetitive microtrauma, plantar fasciitis affects a wide range of people including athletes of all ages [1, 11]. It is particularly common among runners, accounting for about 8 % of all running-related injuries [1]. Plantar fibromatosis, also known as Ledderhose disease, is a relatively rare disorder caused by the hyperproliferation of benign fibromas in the plantar fascia.

General risk factors for plantar fasciitis include increasing age, high body mass index, a sedentary lifestyle, inadequate muscle strength, and limited ankle dorsiflexion or posterior chain tightness [1, 2, 26]. For athletes, excessive training, training errors, running on hard surfaces, poor biomechanics and inflexibility are additional risk factors for plantar fasciitis [1]. In contrast, risk factors for plantar fibromatosis are not well understood. Repetitive trauma has been shown to contribute to plantar fibromatosis in its early proliferative phase [29].

To our knowledge, the genetic basis for plantar fasciitis has not been previously studied. Evidence for a genetic basis of plantar fibromatosis stems from two cytogenetic studies; specifically,

trisomies of chromosomes 8 and 14 were reported in one patient, and a reciprocal translocation was observed in another patient [4, 28]. An unanswered question is whether there may also be genetic differences that affect an individual's inherent risk for plantar fascial disorders.

To begin to address this question, we performed the first genetic association study for plantar fascial disorders and thereby identified two DNA loci that show genome-wide significant associations.

Methods

A genome-wide association screen was performed for plantar fascial disorders using data from the genotyped Genetic Epidemiology Research on Adult Health and Aging (GERA) cohort of the RPGEH. The data generation and data analysis pipeline have been previously described in Roos et al. 2016 [27]. A complete description of the cohort and study design can be found in dbGaP (Study Accession: phs000674.v1.p1).

Our analysis cohort (n = 102,503) includes 59,737 females, 42,958 males, and 66 individuals of unknown sex. Moreover, our analysis cohort is racially diverse, including 83,264 European-White (EUR); 8,560 Latino (LAT); 7,518 East Asian (EAS) and 3,161 African-American (AFR) individuals based on ancestry principle components.

Participants were genotyped at over 650,000 SNPs on four ancestry group-specific Affymetrix Axiom genome-wide arrays optimized for individuals of European (EUR), African-American (AFR), East Asian (EAS), and Latino (LAT) ancestry group [12, 13]. The final number of SNPs that were directly genotyped was 670,572 for EUR; 802,186 for LAT; 708,373 for EAS; and 878,176 for AFR arrays.

Genotypes were pre-phased with Shape-IT v2.r644 (https://mathgen.stats.ox.ac.uk/genetics_software/shapeit/shapeit.html; accessed Feb. 2, 2016) [17]. Then genotypes were imputed by using a cosmopolitan reference panel consisting of all individuals from the 1000 Genomes Project (Mar 2012 release) [16] using IMPUTE2 v2.2.2 (https://mathgen.stats.ox.ac.uk/impute/impute_v2.html; accessed Feb. 2, 2016) [17] with a cutoff of $R^2 > 0.3$ [15]. The quality of the imputed data was previously validated in Jorgenson et al., 2015 [19]. The final number of imputed genetic markers accepted for analysis was 31,085,734. Arrays were merged to create a single file and then broken into smaller genomic chunks in the standard IMPUTE2 probabilistic format with a corresponding PLINK family file. The estimated R^2 metric from IMPUTE2 estimates the correlation between the true and imputed genotype and was used to filter SNPs in our GWA analyses [22].

Determination of genetic ancestry was performed by principal component analysis (PCA), as previously described [27]. These ancestry principal components were used in the GWAS to adjust for genetic ancestry.

▶ **Table 1** Plantar fascial disorder phenotypes classified by ICD and/or CPT codes.

Code[a]	Code Description	N[b]
ICD-728.71	Plantar fasciitis/fibromatosis	21,130
ICD-M72.2	Plantar fascial fibromatosis	1,016
CPT-28060	Fasciectomy, plantar fascia; partial (separate procedure)	10
CPT-28062	Fasciectomy, plantar fascia; radical (separate procedure)	4
CPT-28119	Ostectomy, calcaneus; for spur, with or without plantar fascial release	145
CPT-29893	Endoscopic plantar fasciotomy	27

[a]International Statistical Classification of Diseases and Related Health Problems (ICD-9 or ICD-10) and Current Procedural Terminology (CPT-4) codes extracted from KPNC electronic health records of GERA cohort subjects.

[b]Number of individuals. Some of the 21,624 patients had two or more codes.

Phenotype definition

Plantar fascial disorders were identified in the GERA cohort based on clinical diagnoses and surgical procedures captured in the KPNC Electronic Health Record system (▶ **Table 1**). The vast majority of cases (98 %) had the ICD9 code 728.71 (plantar fasciitis/fibromatosis) that does not allow differentiation of plantar fasciitis from fibromatosis. The surgical procedure (CPT-28119, Ostectomy, calcaneus; for spur, with or without plantar fascial release) was used for a relatively small number of cases, and predominantly includes plantar fascial disorders but also some other types of foot pain. The Electronic Health Record contains reported injuries over the entire lifetime of the patients, including those that occurred prior to enrollment in KPNC as well as those that occurred after the genotyping analysis was performed, if reported by the patient and recorded by the physician.

Genome-wide association and meta-analysis

Genome-wide association analyses of the GERA cohort were performed with a logistic regression model using allele counts for typed and imputed SNPs in an additive genetic model for each of the ancestry groups. The model was adjusted for genetic sex, age at enrollment into the RPGEH cohort, ancestry group using principal components, and variations in genotyping protocol. The final number of SNPs that were analyzed was 8,795,348 for EUR; 9,153,118 for LAT; 8,055,053 for EAS and 14,989,676 for AFR populations. To account for inflation due to population stratification, the genomic control parameter (λ) was calculated: EUR (1.053), LAT (1.018), EAS (1.028), AFR (1.007). Subsequently, p-values were adjusted for λ in each population. Results from each population were combined by inverse-variance, fixed-effects meta-analysis as previously described [27]. SNPs that did not contain data for EUR were removed, because EUR comprises more than 80 % of the cohort. The highest DNA variant that was excluded because it lacked data from the EUR ancestry group was an indel (chr11:22175009:I). This indel had a p-value of 7×10^{-6} that was the 35th highest on the list of DNA variants associated with plantar fascial disorders. The final number of SNPs that was analyzed in the fixed-effects meta-analysis was 9,322,588. To account for multiple hypothesis testing, we set the threshold for statistical significance at $p < 5 \times 10^{-8}$ [14, 23, 24]. Summary statistics for all SNPs from the fixed-effects meta-analysis are available at NIH GRASP: https://grasp.nhlbi.nih.gov/FullResults.aspx.

We examined the level of heterogeneity between ancestry groups using two measures: 1) the I^2 statistic, which measures the

percentage of variability across ancestry groups that is due to heterogeneity, where a lower value indicates more consistent results across ancestry groups, and 2) Cochran's Q statistic, which measures whether observed differences in results between different ancestry groups are due to chance alone, where a low associated p-value indicates heterogeneity [5, 10]. The 95 % confidence interval for I^2 was calculated using the heterogi module for STATA.

Further bioinformatics investigation of the top genome-wide significant loci from the meta-analysis was conducted. The genomic context of each SNP was investigated using RegulomeDB web tools (http://regulomedb.org/; accessed July 1, 2016) [3]. Whether each SNP is an expression quantitative trait locus (eQTL) was queried using the NCBI eQTL Browser (http://www.ncbi.nlm.nih.gov/projects/gap/eqtl/index.cgi; accessed July 1, 2016) and the Genotype-Tissue Expression (GTEx) Portal (http://www.gtexportal.org/home/; accessed July 1, 2016) [7]. Location of SNPs within transcription factor binding sites or DNAse I hypersensitive regions was queried using data from ENCODE (https://www.genome.gov/10005107/encode-project/; accessed July 1, 2016) [6].

Ethical considerations

This study analyzed stored data from RPGEH subjects who consented to genomic testing and use of their genomic data, as well as health data from the KPNC Electronic Health Record, for future research studies. The health and genotype data for the subjects were de-identified. All study procedures were approved by the Kaiser Foundation Research Institute Institutional Review Board. This paper conforms to the ethical standards established by this journal [9].

Results

Study population and genotype information

We performed a logistic regression for SNPs associated with plantar fascial disorders using genotype and medical data from 102,503 individuals that included 21,624 cases and 80,879 controls from each of four ancestry groups: European, Latin-American, African-American and East Asian. A description of the sex and ancestry group of the cases and controls is shown in ▶ **Table 2**. The International Classification of Diseases (ICD) 9 and 10 as well as Current Procedural Terminology (CPT) codes used to identify cases from the electronic medical records are shown in ▶ **Table 1**. Overall, the period prevalence of plantar fascial disorders was 21.1 %.

We combined the data from each of the ancestry groups to generate an overall p-value. ▶ **Fig. 1** shows an overview of the analysis for this genome-wide association study. We compared the observed p-values from the meta-analysis to the distribution of p-values that would be expected by chance in a Q-Q plot (▶ **Fig. 2**). We saw deviation from the null hypothesis for many SNPs with observed p-values below 10^{-5}. This observation indicates that there are more SNPs with low p-values than would be expected by chance; i. e., the deviation from the red line in the upper right portion of ▶ **Fig. 2** is caused by SNPs showing weak associations with plantar fascial disorders.

The p-value for every SNP from the meta-analysis is shown in a Manhattan plot in ▶ **Fig. 3**. To account for multiple hypothesis testing, we set the threshold for statistical significance at $p < 5 \times 10^{-8}$ (indicated by the red line) [14, 23, 24]. There are two independent DNA variants associated with plantar fascial disorders with p-values that are genome-wide significant. The first genetic variant is an indel (chr5:118704153:D) on chromosome 5 with a p-value of 3.04×10^{-8} (▶ **Table 3**). Additionally, there are 14 more SNPs in the same linkage disequilibrium block that are correlated with chr5:118704153:D (using $R^2 > .7$ as a cutoff). As expected, these 14 linked SNPs also show associations with plantar fascial disorders, albeit at weaker levels (**Supplemental Table 1**). Hence, these 15 genetic variants on chromosome 5 represent one association signal from a linkage disequilibium block, with chr5:118704153:D being the most significant.

The second DNA variant, rs62051384, is on chromosome 16 and is associated with plantar fascial disorders with a p-value of 2.47×10^{-8}. This SNP is in a region that contains 43 linked SNPs that also show associations with plantar fascial disorders at weaker levels (**Supplemental Table 1**). Hereafter, we refer to the two regions on chromosomes 5 and 16 using the sentinel genetic variants chr5:118704153:D and rs62051384.

Almost none of the DNA variants in the loci on chromosomes 5 or 16 were directly genotyped on the Affymetrix chips, but rather their genotype data was imputed (**Supplemental Table 1**). The R^2 value was 0.93 for chr5:118704153:D and 0.99 for rs62051384. The R^2 values for the other linked SNPs on chromosomes 5 and 16 were similarly high, indicating that their genotypes were determined fairly accurately using imputation (**Supplemental Table 1**).

For chr5:118704153:D, the TGC allele is associated with decreased risk and the T allele is associated with increased risk for plantar fascial disorders. The frequency of the protective allele (TGC) ranged between 13.6 % in the African-American controls to 40.7 % in the East Asian controls. For rs62051384, T is the risk allele and C is the protective allele for plantar fascial disorders. The frequency of the risk allele (T) ranged between 33.2 % in the African-American controls to 47.5 % in the Latin-American controls. The overall allelic odds ratios for chr5:118704153:D and rs62051384 were 0.93 (95 % CI = 0.91–0.96) and 1.07 (95% CI = 1.04–1.09), respectively.(▶ **Table 3**).

For chr5:118704153:D, individuals that carried one (genotype TGC/T) protective allele had 5.5 % decreased chance of plantar fascial disorders compared to individuals lacking a protective allele (genotype T/T). Individuals that contained two (genotype TGC/TGC) protective alleles had a 9.7 % decreased chance of plantar fascial disorders (▶ **Table 4**). For rs62051384, individuals carrying two (genotype T/T) or one (genotype T/C) risk alleles had 10.8 % and 3.9 % increased risk compared to individuals lacking a risk allele (genotype C/C), respectively (▶ **Table 4**).

The GWAS results were analyzed to determine whether the association with plantar fascial disorders for either chr5:118704153:D or rs62051384 was stronger in some ancestry groups than in others, a phenomenon known as heterogeneity [18]. ▶ **Table 5** shows the p-values and odds ratios for these two genetic variants for each of the ancestry groups. For both genetic variants, the smallest p-value was observed for the European population, which is expected because 81 % of the cohort was European. The odds ratios for each ancestry group were in the same direction and of similar magnitude. Using I^2 and Cochran's Q to assess heterogeneity, we saw no evidence of heterogeneity for either genetic variant (▶ **Table 5**).

▶ **Table 2** Demographic factors used in genome-wide association analyses of plantar fascial disorders.

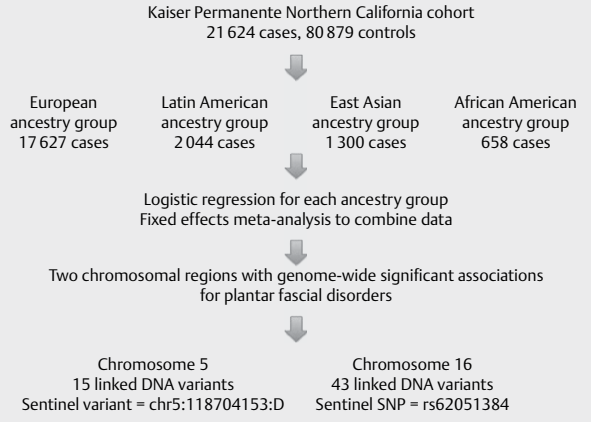
	Cases[a]	Controls	Overall
Subjects (%)	21,624 (21.1%)	80,879 (78.9%)	102,503
Sex (%)[b]			
Male	8,256 (19.2%)	34,702 (80.8%)	42,958
Female	13,254 (22.3%)	46,125 (77.7%)	59,379
Undetermined	11 (1.5%)	55 (98.5%)	66
Ancestry Group (%)[c]			
European	17,627 (21.2%)	65,637 (78.8%)	83,264
Latin American	2,044 (23.9%)	6,516 (76.1%)	8,560
East Asian	1,300(17.3%)	6,218 (72.7%)	7,518
African American	658 (20.8%)	2,508 (79.2%)	3,161
Age[d]	63.1 (62.9-63.2)	62.4 (62.3-62.5)	62.6 (62.6-62.7)

[a] Cases with plantar fascial disorders

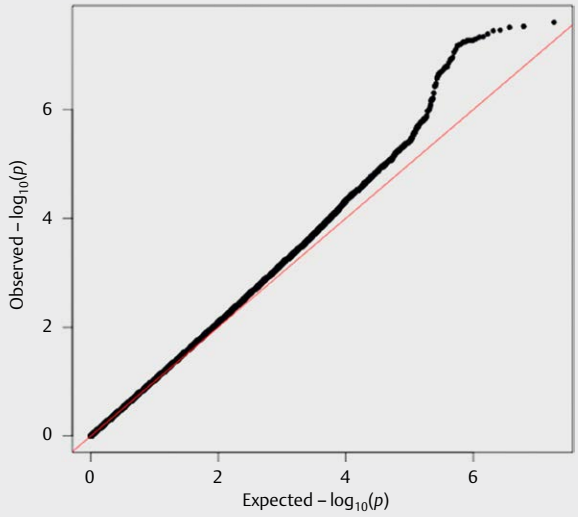
[b] Sex/gender as determined by an individual's genetic data, reported as the number and percentage of total

[c] Ancestry groups as determined by principle components analysis

[d] Average age at subject enrollment (95% CI)



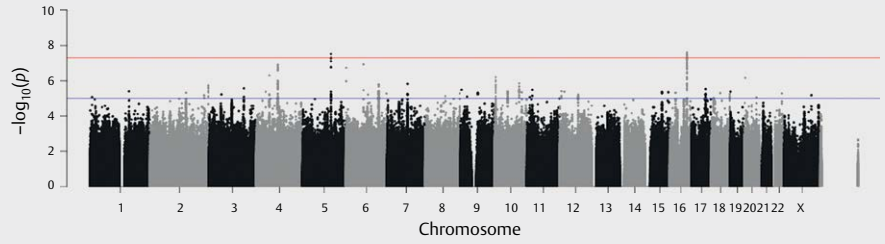
▶ **Fig. 1** Overview of the workflow for the genome-wide association study for plantar fascial disorders.



▶ **Fig. 2** Quantile-quantile plot for genome-wide association analyses of plantar fascial disorders. The expected versus observed log-transformed values for the 9,322,588 p-values are graphed. The observed p-values (black dots) are plotted on the y-axis and the p-values expected by chance (red line) are plotted on the x-axis.

The indel chr5:118704153:D is located within an intron of TNFAIP8 (TNF alpha-induced protein 8) (▶ **Fig. 3**). TNFAIP8 encodes a protein with a death effector domain and its expression is induced by the cytokine TNF alpha in thymocytes, which protects them from undergoing apoptosis [8, 21]. We searched for a mechanism whereby chr5:118704153:D or a linked SNP might affect the activity of TNFAIP8, possibly accounting for the association with plantar fascial disorders. Any one of the 15 genetic variants in the locus on chromosome 5 might be responsible for the association affecting risk for plantar fascial disorders by affecting one or more of the nearby genes (i. e., a weak mutation), with the remaining genetic variants possibly having no effect on either gene function or phenotype (i. e., neutral polymorphism). None of the 15 genetic variants are in coding regions for TNFAIP8. All 15 of the SNPs are associated with changes in expression of TNFAIP8 (i. e., expression quantitative trait loci) [7]. For all genetic variants, the minor allele is associated with higher expression of TNFAIP8; e. g., the TGC allele of chr5:118704153:D is protective for plantar fascial disorders and also associated with higher expression of TNFAIP8 than the T allele. Among the SNPs in this linkage disequilibrium block, rs1509142 has the best evidence for being a causal mutation that directly affects expression of TNFAIP8 [3]. The rs1509142 SNP lies within a region bound by three transcription factors: MAZ, NFIC and NFKB1 [6]. Allelic variation at rs1509142 might affect binding of one or more of these transcription factors, thereby altering expression of TNFAIP8.

The rs62051384 SNP is located within an intron of WWP2 (WW domain-containing protein 2) gene, which encodes a ubiquitin-protein ligase involved in proteasomal degradation (▶ **Fig. 4**) [25].



▶ **Fig. 3** Manhattan plot for genome-wide association analyses of plantar fascial disorders. The -$\log_{10}$ p-values for association with plantar fascial disorders for SNPs from the meta-analysis are plotted by genomic position with chromosome number listed across the bottom. The y-axis shows the -$\log_{10}$ p-value for association with plantar fascial disorders. The blue line represents suggestive genome-wide significance ($p < 1 \times 10^{-5}$) and the red line represents genome-wide significance ($p < 5 \times 10^{-8}$).

▶ **Table 3** Genome-wide significant associations for plantar fascial disorders.

SNP	Gene	EA[a]	EAF[b]	P-value[c]	OR (95 % CI)[d]
chr5:118704153:D	TNFAIP8	TGC	0.322	3.04×10^{-8}	0.93 (0.91-0.96)
rs62051384	WWP2	T	0.373	2.47×10^{-8}	1.07 (1.04-1.09)
[a] Effect allele. The other allele is listed in Supplemental Table 3					
[b] Effect allele frequency in the control population					
[c] P-value from fixed-effects meta-analysis. The cut-off for genome-wide significant association was $p = 5 \times 10^{-8}$					
[d] Allelic odds ratio with 95 % confidence interval					

The linkage block on chromosome 16 containing rs62051384 also contains 42 tightly-linked genetic variants ($R^2 > 0.7$)(**Supplemental Table 1**). Similarly to the tightly-linked variants in the chromosome 5 block, we investigated whether any of these 43 variants affected the coding sequence or expression levels of nearby genes. None of these 43 SNPs are located in a protein-coding region. Among the SNPs located in this region, rs6499257 is a good candidate for causing changes in expression in WWP2. The G allele of the rs6499257 SNP is associated with higher expression of WWP2 and increased risk for plantar fascial disorders [7]. ChIP-seq data from the ENCODE project indicate that rs6499257 lies within a DNAse I hypersensitive region, which marks regions bound by transcription factors [6] (▶ **Fig. 5**).

Discussion

Plantar fascial disorders include both plantar fasciitis and plantar fibromatosis. Plantar fasciitis is the most common cause of heel pain in adults, and afflicts athletes participating in sports involving running or jumping [2]. Plantar fibromatosis, on the other hand, is a relatively rare hyperproliferative disorder of the plantar fascia.

Very little is known about genetic risks in the etiology of either plantar fasciitis or fibromatosis. Cytogenetic studies have identified a chromosome duplication and a reciprocal translocation in two cases of plantar fibromatosis, suggesting that these chromosomal abnormalities contributed to the formation of the fibroma [4, 28].

By obtaining access to large-scale genotype and phenotype data from the RPGEH, we were able to find the first evidence for specific genetic polymorphisms associated with these plantar fascial disorders. The data contained information from 102,503 individuals of whom 21,624 were treated for a plantar fascial disorder. The data from this study indicate that genetic differences appear to affect an individual's intrinsic risk for plantar fascial disorders.

Our results showed genome-wide significant associations for chr5:118704153:D and rs62051384 with plantar fasciitis or fibromatosis. As the ICD-9 (728.71) and ICD-10 (M72.2) codes include both diagnoses of plantar fasciitis and fibromatosis, we cannot discern whether the genetic variants are associated with fasciitis, fibromatosis or both. The reason we analyzed plantar fasciitis and plantar fibromatosis together was that they are combined in the same ICD-9 code (ICD-728.71, Plantar fasciitis/fibromatosis). Since we do not have access to the patients themselves, the coding set up by the International Statistical Classification of Diseases and Related Health Problems committees constrained us to do the analysis on the combined fasciitis/fibromatosis phenotypes. We do not know if the etiology or underlying mechanisms are shared between plantar fasciitis and fibromatosis. If they were to share a common underlying genetic mechanism, then combining the two injuries would improve the statistical power of the analysis. If they were to have different underlying genetic mechanisms, then our analysis would most likely reveal DNA variants associated with plantar fasciitis, which make up the majority of the cases coded with ICD-728.71.

Our cohort included people regardless of whether or not they participated in a sport. It is unknown whether the statistical association of chr5:118704153:D and rs62051384 with plantar fascial disorders derives predominantly from the subset of the population that is active in sports. It is also possible that these genetic variants act by increasing a known risk factor for plantar fascial disorders. High body mass index is one of the risk factors for plantar fasciitis, and the genetic basis for high body mass index has been studied

extensively [3]. However, neither variant was found to be associated with increased body mass index in published studies, so it is unlikely that these two DNA variants increase the risk for plantar fascial disorders by increasing body mass index. Other risk factors for plantar fasciitis include inadequate muscle strength, limited ankle dorsiflexion and posterior chain tightness [1, 2, 26]. The genetic basis for these risk factors is poorly understood, so it is possible that the two DNA variants identified in this study act first by increasing the risk for one of these factors and then by increasing the risk for plantar fasciitis. Another explanation is that these polymorphisms are associated with an increased propensity to exercise and perform sports that have a high risk of developing plantar fasciitis.

The indel chr5:118704153:D is located within TNFAIP8 on chromosome 5. This gene encodes a protein induced by TNF alpha, and is involved in protecting thymocytes from undergoing apoptosis [21]. The rs62051384 SNP is located within WWP2, which is involved in proteasomal degradation [25]. Currently, the mechanistic link between the biochemical functions for either of these proteins and the onset of plantar fasciitis or fibromatosis is unclear.

None of the genetic variants in the linkage blocks on chromosome 5 or 16 that show an association with plantar fasciitis or fibromatosis are located in coding regions. However, chr5:118704153:D and rs62051384 (as well as other tightly-linked genetic variants in the same linkage disequilibrium blocks) are associated with variation in expression of TNFAIP8 and WWP2, respectively. It is not yet possible to discern which specific variant within these blocks is responsible for variation in expression of these two genes. Currently, data from the ENCODE project suggest that rs1509142 (on chromosome 5) is the most likely SNP for being directly responsible for affecting expression of TNFAIP8, because this SNP is located in the DNA region bound by three transcription factors (MAZ, NFIC and NFKB1). For the chromosome 16 locus, rs6499257 is the most likely SNP for being directly responsible for affecting expression of WWP2. This SNP is located in a DNAse I hypersensitive region, which is typically bound by transcription factors [6]. One possibility is that allelic variation at these two loci might affect binding of a transcription factor, thereby affecting expression of linked genes (e. g., TNFAIP8 or WWP2).

What is the clinical impact of these genetic markers on plantar fascial disorders? For chr5:118704153:D, about 9 % of individuals were homozygous for the protective allele (TGC/TGC) in our study population and had a 9.7 % decreased risk for plantar fascial disorders compared to the 40 % of individuals that were homozygous for the risk allele (T/T). For rs62051384, 14 % were homozygous for the risk allele (T/T) and had a 10.8 % increased risk compared to the 38 % of individuals that were homozygous for the protective allele (C/C). These genetic markers help predict the risk for a common injury affecting about 21 % of the general population. Even a small reduction in overall risk could benefit many people because the injury is so common. For the general population, the effect sizes from these two SNPs (± 10 % risk) are relatively small and would not warrant lifestyle changes. For elite athletes, however, a 10 % change in risk for plantar fascial disorders may warrant attention with regard to training regimen.

▶ **Table 4** Genotype distributions for chr5:118704153:D and rs62051384.

chr5:118704153:D	TGC/TGC[a]	TGC/T[a]	T/T[a]
Cases	1,838	7,751	9,015
Controls	7,537	30,139	32,519
Overall	9,375	37,890	41,534
Risk for plantar fascial disorders	0.196	0.205	0.217
Relative risk for plantar fascial disorders[b]	0.903	0.945	1.000
rs62051384	**T/T**	**T/C**	**C/C**
Cases	3,275	9,775	7,968
Controls	11,288	36,450	31,276
Overall	14,563	46,225	39,244
Risk for plantar fascial disorders	0.225	0.211	0.203
Relative risk for plantar fascial disorders[a]	1.108	1.039	1.000

[a]Number of individuals with that genotype

[b]Risk for plantar fascial disorders given that genotype compared to the overall risk

▶ **Table 5** Association statistics for chr5:118704153:D and rs62051384 in individual ancestry groups.

Ancestry Group	SNP	EA[a]	P-value[c]	OR (95 % CI)[d]	I^2 (95 % CI)[d]	Q[e]
EUR	chr5:118704153:D	TGC	3.52×10^{-7}	0.93 (0.9-.96)	0 (0-85)	0.80
LAT	chr5:118704153:D	TGC	.034	0.91 (0.84-0.99)		
AFR	chr5:118704153:D	TGC	.37	0.91 (0.74-1.12)		
EAS	chr5:118704153:D	TGC	.50	0.97 (0.88-1.06)		
EUR	rs62051384	T	1.3×10^{-6}	1.06 (1.04-1.09)	58 (0-85)	0.65
LAT	rs62051384	T	.89	1.00 (0.94-1.08)		
AFR	rs62051384	T	9.1×10^{-3}	1.20 (1.04-1.36)		
EAS	rs62051384	T	4.6×10^{-3}	1.13 (1.04-1.23)		

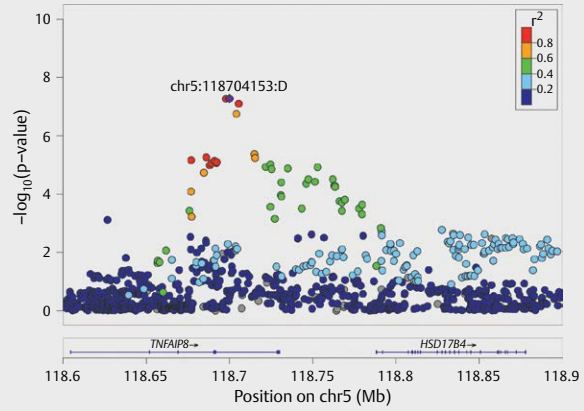
[a]Effect allele

[b]P-value adjusted for lambda genomic inflation factor

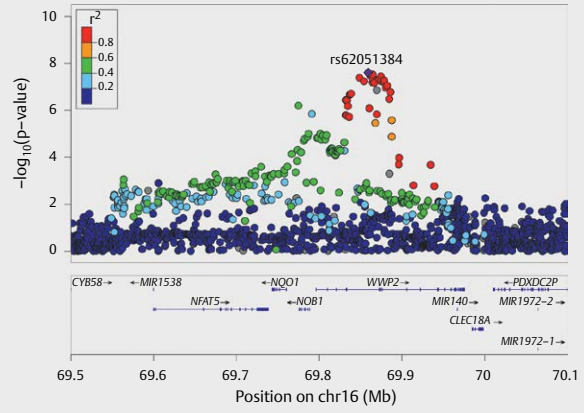
[c]Odds ratio (95 % confidence interval)

[d]Percentage of variability between ancestry groups that is due to heterogeneity (95 % confidence interval)

[e]Cochran's Q, p-value that the association is different between ancestry groups



▶ **Fig. 4** Regional association plot for chr5:118704153:D with plantar fascial disorders. Tested SNPs are arranged by genomic position on chromosome 5 (x-axis) in a 300 kb window around the lead SNP chr5:118704153:D (purple diamond). The y-axis indicates -$\log_{10}$ p-values for association with plantar fascial disorders for each SNP. chr5:118704153:D is located in an intron of TNFAIP8. The color of dots of the flanking SNPs indicates their linkage disequilibrium (R^2) with the lead SNP as indicated in the heat map color key.



▶ **Fig. 5** Regional association plot for rs62051384 with plantar fascial disorders. Tested SNPs are arranged by genomic position on chromosome 16 (x-axis) in a 600 kb window around the lead SNP rs62051384 (purple diamond). The y-axis indicates -$\log_{10}$ p-values for association with plantar fascial disorders for each SNP. rs62051384 is located within the gene WWP2. The color of dots of the flanking SNPs indicates their linkage disequilibrium (R^2) with the lead SNP as indicated in the heat map color key.

There are several limitations to this study. First, the phenotypes were defined from codes contained in the electronic health records, which may be inaccurate. The accurate diagnosis of plantar fasciitis is challenging due to the regional anatomy and the fact that it is generally a clinical diagnosis obtained from history and physical findings [20]. For example, in addition to plantar fasciitis and fibromatosis, it is possible that ICD and CPT codes listed in ▶ **Table 1** might also include cases of heel pain due to neurological impingement or heel pad disorders. Furthermore, we are unable to differentiate plantar fasciitis from fibromatosis by ICD code (though plantar fasciitis is far more common than plantar fibromatosis). Second, the number of individuals of Latin-American, African-American and Asian ethnicity was relatively small, and hence the association results for these results are weaker than those from the European group. Third, since this study was performed using the general population rather than athletes, the results reported here might be different than the results for a study using only participants that are active in sports.

In the future, it will be important to replicate the gene association results with plantar fascial disorders in an independent cohort. It will also be interesting to perform the analysis on a population of athletes competing in sports with high rates of plantar fasciitis or fibromatosis, such as long-distance running or triathlons. The results from these studies may reveal that certain genetic polymorphisms, such as chr5:118704153:D or rs62051384, could be used as diagnostic markers to predict which athletes are at higher risk for injury. Preventative measures could then be taken to alleviate that risk, thereby reducing the overall incidence. Finally, additional studies are warranted to begin to illuminate the underlying biological mechanism for the association of variations near TNFAIP8 and WWP2 with plantar fasciitis or fibromatosis.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] Beeson P. Plantar fasciopathy: Revisiting the risk factors. Foot Ankle Surg 2014; 20: 160–165

[2] Bolivar YA, Munuera PV, Padillo JP. Relationship between tightness of the posterior muscles of the lower limb and plantar fasciitis. Foot Ankle Int 2013; 34: 42–48

[3] Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M, Karczewski KJ, Park J, Hitz BC, Weng S, Cherry JM, Snyder M. Annotation of functional variation in personal genomes using RegulomeDB. Genome Res 2012; 22: 1790–1797

[4] Breiner JA, Nelson M, Bredthauer BD, Neff JR, Bridge JA. Trisomy 8 and trisomy 14 in plantar fibromatosis. Cancer Genet Cytogenet 1999; 108: 176–177

[5] Cochran WG. The combination of estimates from different experiments. Biometrics 1954; 10: 101–129

[6] Consortium EP. An integrated encyclopedia of DNA elements in the human genome. Nature 2012; 489: 57–74

[7] Consortium GT. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. Science 2015; 348: 648–660

[8] Cox A, Dunning AM, Garcia-Closas M, Balasubramanian S, Reed MW, Pooley KA, Scollen S, Baynes C, Ponder BA, Chanock S, Lissowska J, Brinton L, Peplonska B, Southey MC, Hopper JL, McCredie MR, Giles GG, Fletcher O, Johnson N, dos Santos Silva I, Gibson L, Bojesen SE, Nordestgaard BG, Axelsson CK, Torres D, Hamann U, Justenhoven C, Brauch H, Chang-Claude J, Kropp S, Risch A, Wang-Gohrke S, Schurmann P, Bogdanova N, Dork T, Fagerholm R, Aaltonen K, Blomqvist C, Nevanlinna H, Seal S, Renwick A, Stratton MR, Rahman N, Sangrajrang S, Hughes D, Odefrey F, Brennan P, Spurdle AB, Chenevix-Trench G, Kathleen Cunningham Foundation Consortium for Research

into Familial Breast C.Beesley J, Mannermaa A, Hartikainen J, Kataja V, Kosma VM, Couch FJ, Olson JE, Goode EL, Broeks A, Schmidt MK, Hogervorst FB, Van't Veer LJ, Kang D, Yoo KY, Noh DY, Ahn SH, Wedren S, Hall P, Low YL, Liu J, Milne RL, Ribas G, Gonzalez-Neira A, Benitez J, Sigurdson AJ, Stredrick DL, Alexander BH, Struewing JP, Pharoah PD, Easton DF.Breast Cancer Association C. A common coding variant in CASP8 is associated with breast cancer risk. Nat Genet 2007; 39: 352–358

[9] Harriss DJ, Macsween A, Atkinson G. Standards for Ethics in Sport and Exercise Science Research: 2018 Update. Int J Sports Med 2017; 38: 1126–1131

[10] Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ 2003; 327: 557–560

[11] Hill CL, Gill TK, Menz HB, Taylor AW. Prevalence and correlates of foot pain in a population-based study: The North West Adelaide health study. J Foot Ankle Res 2008; 1: 2

[12] Hoffmann TJ, Kvale MN, Hesselson SE, Zhan Y, Aquino C, Cao Y, Cawley S, Chung E, Connell S, Eshragh J, Ewing M, Gollub J, Henderson M, Hubbell E, Iribarren C, Kaufman J, Lao RZ, Lu Y, Ludwig D, Mathauda GK, McGuire W, Mei G, Miles S, Purdy MM, Quesenberry C, Ranatunga D, Rowell S, Sadler M, Shapero MH, Shen L, Shenoy TR, Smethurst D, Van den Eeden SK, Walter L, Wan E, Wearley R, Webster T, Wen CC, Weng L, Whitmer RA, Williams A, Wong SC, Zau C, Finn A, Schaefer C, Kwok PY, Risch N. Next generation genome-wide association tool: design and coverage of a high-throughput European-optimized SNP array. Genomics 2011; 98: 79–89

[13] Hoffmann TJ, Zhan Y, Kvale MN, Hesselson SE, Gollub J, Iribarren C, Lu Y, Mei G, Purdy MM, Quesenberry C, Rowell S, Shapero MH, Smethurst D, Somkin CP, Van den Eeden SK, Walter L, Webster T, Whitmer RA, Finn A, Schaefer C, Kwok PY, Risch N. Design and coverage of high throughput genotyping arrays optimized for individuals of East Asian, African American, and Latino race/ethnicity using imputation and a novel hybrid SNP selection algorithm. Genomics 2011; 98: 422–430

[14] Hoggart CJ, Clark TG, De Iorio M, Whittaker JC, Balding DJ. Genome-wide significance for dense SNP and resequencing data. Genet Epidemiol 2008; 32: 179–185

[15] Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. Nat Genet 2012; 44: 955–959

[16] Howie B, Marchini J, Stephens M. Genotype imputation with thousands of genomes. G3 (Bethesda) 2011; 1: 457–470

[17] Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet 2009; 5: e1000529

[18] Ioannidis JPA, Patsopoulos NA, Evangelou E. Heterogeneity in meta-analyses of genome-wide association investigations. PLoS One 2007; 2: e841

[19] Jorgenson E, Makki N, Shen L, Chen DC, Tian C, Eckalbar WL, Hinds D, Ahituv N, Avins A. A genome-wide association study identifies four novel susceptibility loci underlying inguinal hernia. Nat Commun 2015; 6: 10130

[20] Lareau CR, Sawyer GA, Wang JH, DiGiovanni CW. Plantar and medial heel pain: diagnosis and management. J Am Acad Orthop Surg 2014; 22: 372–380

[21] Lou Y, Liu S. The TIPE (TNFAIP8) family in inflammation, immunity, and cancer. Mol Immunol 2011; 49: 4–7

[22] Marchini J, Howie B. Genotype imputation for genome-wide association studies. Nat Rev Genet 2010; 11: 499–511

[23] Panagiotou OA, Ioannidis JPA.Genome-Wide Significance Project. What should the genome-wide significance threshold be? Empirical replication of borderline genetic associations. Int J Epidemiol 2012; 41: 273–286

[24] Pe'er I, Yelensky R, Altshuler D, Daly MJ. Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. Genet Epidemiol 2008; 32: 381–385

[25] Pirozzi G, McConnell SJ, Uveges AJ, Carter JM, Sparks AB, Kay BK, Fowlkes DM. Identification of novel human WW domain-containing proteins by cloning of ligand targets. J Biol Chem 1997; 272: 14611–14616

[26] Riddle DL, Pulisic M, Pidcoe P, Johnson RE. Risk factors for Plantar fasciitis: a matched case-control study. J Bone Joint Surg Am 2003; 85-a: 872–877

[27] Roos TR, Roos AK, Avins AL, Ahmed MA, Kleimeyer JP, Frederidson M, Ioannidis JPA, Dragoo JL, Kim SK. Genome-wide association study identifies a locus associated with rotator cuff injury. PLoS One 2017; 12: e0189317–1746

[28] Sawyer JR, Sammartino G, Gokden N, Nicholas RW. A clonal reciprocal t(2;7)(p13;p13) in plantar fibromatosis. Cancer Genet Cytogenet 2005; 158: 67–69

[29] Veith NT, Tschernig T, Histing T, Madry H. Plantar fibromatosis--topical review. Foot Ankle Int 2013; 34: 1742–1746

Supplementary Material

▶ **Supplemental Table 1**

DNA variants on Chromosome 5								
Variant	**EA**	**A2**	**P-value**	**OR (95 % CI)**	**R2 to sentinel chr5:118704153:D**	**R2 from imputation**	**BP**	**Distance to sentinel chr5:118704153:D**
rs250303	G	A	6.72E-06	0.95 (0.93-0.97)	0.76	0.99	118677093	-27060
rs250309	T	A	1.85E-05	0.95 (0.93-0.97)	0.77	0.98	118684602	-19551
rs192225	T	A	5.50E-06	0.95 (0.93-0.97)	0.77	0.98	118686083	-18070
rs6874848	T	A	1.03E-05	0.95 (0.93-0.97)	0.79	0.99	118688115	-16038
rs1155268	G	A	8.84E-06	0.95 (0.93-0.97)	0.79	0.99	118689221	-14932
rs3813309	A	G	7.09E-06	0.95 (0.93-0.97)	0.8	0.99	118690998	-13155
rs1509141	G	C	8.50E-06	0.95 (0.93-0.97)	0.8	0.99	118692218	-11935
rs1509142	A	G	8.04E-06	0.95 (0.93-0.97)	0.8	0.99	118692465	-11688
rs6595186	A	G	5.33E-08	0.94 (0.92-0.96)	0.89	0.99	118697849	-6304
rs4895369	A	G	5.30E-08	0.94 (0.92-0.96)	0.89	1	118699951	-4202
chr5:118704153:D	TGC	T	**3.04E-08**	0.93 (0.91-0.95)	1	0.93	118704153	0
rs62375118	C	T	1.76E-07	0.93 (0.9-0.96)	0.87	0.87	118704155	2
rs32654	C	G	5.00E-08	0.94 (0.92-0.96)	0.9	0.99	118704610	457
rs32652	T	G	7.79E-08	0.94 (0.92-0.96)	0.91	0.97	118705545	1392
chr5:118707313:D	GA	G	1.61E-07	0.94 (0.92-0.96)	0.89	0.98	118707313	3160
SNPs on Chromosome 16								
Variant	**EA**	**A2**	**P-value**	**OR (95 % CI)**	**R2 to sentinel rs62051384**	**R2 from imputation**	**BP**	**Distance to sentinel rs62051384**
rs4275849	A	G	1.59E-06	1.06 (1.04-1.08)	0.93	0.98	69832105	-27077
rs4341733	C	T	3.66E-07	1.06 (1.04-1.08)	0.98	0.98	69832247	-26935
rs8063219	G	T	1.66E-06	1.06 (1.04-1.08)	0.93	0.98	69832912	-26270
rs6499257	T	G	3.70E-07	1.06 (1.04-1.08)	0.98	0.98	69833663	-25519
chr16:69834078:D	T	TTTTC	1.55E-05	1.05 (1.03-1.07)	0.94	0.96	69834078	-25104
chr16:69834079:D	T	TTTC	1.97E-05	1.05 (1.03-1.07)	0.94	0.96	69834079	-25103
rs111335581	A	G	6.51E-07	1.06 (1.04-1.08)	0.88	0.95	69834529	-24653
rs11075738	C	T	1.93E-06	1.06 (1.04-1.08)	0.93	0.98	69836125	-23057
rs11643817	T	G	2.16E-07	1.06 (1.04-1.08)	0.98	0.98	69836805	-22377
rs4985418	T	C	2.05E-07	1.06 (1.04-1.08)	0.98	0.98	69837465	-21717
rs8044920	T	C	1.94E-07	1.06 (1.04-1.08)	0.98	0.98	69838676	-20506
rs7193221	G	C	**4.06E-08**	1.07 (1.05-1.09)	0.94	0.99	69848721	-10461
rs60941540	T	A	5.72E-08	1.06 (1.04-1.08)	0.99	0.99	69854029	-5153
rs62051384	T	C	**2.47E-08**	1.07 (1.05-1.09)	1	0.99	69859182	0
rs62051387	T	C	8.18E-07	1.06 (1.04-1.08)	0.95	0.97	69860655	1473

▶ **Supplemental Table 1**

rs62051389	A	G	**3.44E-08**	1.06 (1.04-1.08)	1	0.99	69862249	3067
rs11075739	T	C	**6.37E-08**	1.06 (1.04-1.08)	1	0.99	69864234	5052
rs7194806	A	T	**4.55E-08**	1.06 (1.04-1.08)	1	0.99	69864491	5309
rs6499260	A	G	**2.93E-08**	1.06 (1.04-1.08)	1	0.99	69864675	5493
rs9630631	C	T	6.64E-08	1.06 (1.04-1.08)	1	0.99	69866727	7545
rs4985445	G	A	3.45E-06	1.06 (1.04-1.08)	0.71	0.99	69867835	8653
rs4985374	C	A	1.46E-06	1.06 (1.04-1.08)	0.97	0.96	69869725	10543
rs12447095	G	T	1.38E-07	1.06 (1.04-1.08)	0.99	0.98	69869763	10581
rs7206585	A	G	**4.63E-08**	1.06 (1.04-1.08)	0.99	0.99	69872404	13222
rs8059645	T	G	**3.52E-08**	1.06 (1.04-1.08)	0.99	0.99	69874709	15527
rs7202223	G	A	5.71E-08	1.06 (1.04-1.08)	0.99	0.99	69875191	16009
rs11641899	A	G	6.21E-08	1.06 (1.04-1.08)	0.99	0.99	69877930	18748
rs2173714	G	C	5.31E-08	1.06 (1.04-1.08)	0.99	0.99	69878534	19352
chr16:69879812:I	GC	G	1.83E-05	1.06 (1.04-1.08)	0.93	0.92	69879812	20630
chr16:69879813:D	C	CAT	8.44E-05	1.05 (1.02-1.08)	0.92	0.91	69879813	20631
rs4985446	C	T	1.09E-07	1.06 (1.04-1.08)	0.99	0.99	69880293	21111
rs34378064	A	G	8.81E-08	1.06 (1.04-1.08)	0.99	1	69882083	22901
rs12447914	C	G	3.28E-07	1.06 (1.04-1.08)	0.99	1	69884723	25541
rs68039170	C	T	1.65E-07	1.06 (1.04-1.08)	0.99	1	69886089	26907
chr16:69886574:D	T	TTGTTCA	2.23E-07	1.06 (1.04-1.08)	0.98	0.99	69886574	27392
rs3790086	G	C	2.64E-06	1.06 (1.04-1.08)	0.74	1	69887707	28525
rs11645565	A	G	0.0002001	1.05 (1.03-1.07)	0.79	0.99	69896248	37066
rs7196917	G	A	0.0001043	1.05 (1.03-1.07)	0.8	1	69896527	37345
rs8052428	T	C	0.001567	1.04 (1.02-1.06)	0.85	0.92	69913996	54814
rs7200005			NA		0.76	0.96	69929137	69955
rs11075745	T	C	0.0002079	1.05 (1.03-1.07)	0.75	0.95	69934672	75490
rs8043925	T	C	0.001679	1.04 (1.02-1.06)	0.73	0.94	69939190	80008
rs11075747			NA		0.74	0.89	69940361	81179

EA: Effect Allele

P-value: p-value from fixed-effects meta-analysis. Bold indicates genome-wide significance

OR (95% CI): Allelic odds ratio (95% confidence interval)

BP: position (HG18)