An association between polymorphisms within the *APOE* gene and concussion aetiology in rugby union players

Shameemah Abrahams[a], Sarah McFie[a], Jon Patricios[b,c,d], Jason Suter[e], Michael Posthumus[a], Alison V. September[a,*]

[a] Division of Exercise Science and Sports Medicine, Department of Human Biology, Faculty of Health Science, University of Cape Town, South Africa

[b] Morningside Sports Medicine Clinic, South Africa

[c] Section of Sports Medicine, University of Pretoria, South Africa

[d] Department of Emergency Medicine, University of the Witwatersrand, South Africa

[e] Sports Science and Exercise Medicine Clinic, South Africa

∗ Corresponding author. E-mail address: alison.september@uct.ac.za (A.V. September).

Abstract

Objectives: Concussion refers to changes in neurological function due to biomechanical forces transmitted to the head. The *APOE* ε4 allele is associated with brain injury severity. The objective was to determine if *APOE* gene variants are associated with concussion history and severity in rugby players.

Design: In total, 128 non-concussed controls and 160 previously concussed participants (all cases N=160; diagnosed N=139) were recruited from high school (junior, N=121), club (N=116) and professional rugby teams (N=51).

Methods: Participants were genotyped for rs405509 (G>T), rs429358 (T>C) and rs7412 (C>T) *APOE* variants. Statistical analyses were performed using the R environment.

Results: The rs405509 TT genotype was over-represented in controls compared to all cases (P=0.043; control: 29%, all cases: 18%; odds ratio: 0.55, 95% confidence interval 0.31 – 0.98). The *APOE*-ε isoform frequencies were not significantly different between groups (P>0.05). Additionally, the inferred *APOE* (rs405509-ε2/ε3/ε4) T-ε3 haplotype was over-represented in controls (41%) compared to diagnosed (32%, P=0.042). The G-ε3 haplotype was under-represented in controls (36%) compared to all cases (44%, P=0.019) and diagnosed (44%, P=0.021). The TT genotype was significantly associated with rapid recovery (P=0.048, < 1 week: 51%, N=70, ≥ 1 week: 36%, N=29; odds ratio: 0.55, 95% confidence interval 0.30 – 1.01). **Conclusions**: These findings support the further elucidation of the *APOE* gene or closely-related genes in concussion aetiology. Although similar preliminary results were found when juniors were separately analysed, the under-powered sample size for junior subgroup requires future investigation in larger cohorts of junior-level athletes.

Key words: apolipoprotein E, brain injuries, genetic association studies, promoter region, genetic polymorphism, apoptosis

1. Introduction

Concussion is defined as a direct or indirect transmission of biomechanical forces to the head or neck, resulting in neurological deficits including physical, cognitive, or emotional [1,2]. These deficits may be attributed to ionic disturbances, the dysregulation of cerebral protein metabolism or elevated tissue damage markers with resultant neuronal cell death [3]. Concussions are common in contact sports with an incidence of 4.4 to 8.0 injuries/1000 player hours reported in semi-professional rugby [4]. Several risk factors for concussion have been proposed including previous history, game play, age, gender and genetics [5]. Clinically, certain vulnerable groups have also been highlighted, particularly children and adolescents [6]. Young athletes may represent an extreme phenotype for concussion because they suffer prolonged recovery from symptoms and neurocognitive deficits with recovery lasting 10 – 14 days (double the symptom duration commonly observed) [7,8]. Additionally, youth athletes seem to

experience severe complications following concussion (e.g. fatal cerebral swelling or 'second impact syndrome') [9]. The potential risk of debilitating concussion outcomes for youth athletes emphasise the importance to investigate the genetic role on concussion in youth athletes. Although there is limited research on the genetic contribution to concussion risk, genetic variants have been associated with concussion susceptibility [10,11] and severity [12,13]. Concussion severity can be determined from the clinical recovery decision which is based on symptom resolution and neurocognitive improvement [16,17].

A promoter polymorphism (rs405509, -219 G>T) within the *apolipoprotein E* gene (*APOE*), shown to alter *APOE* mRNA expression [14], and two non-synonymous *APOE* polymorphisms (rs429358: T>C, p.Arg112Cys, rs7412: C>T, p.Arg158Cys), which result in one of four apo-E isoforms (E2, E3, E4, rare E3r) [15], have previously been associated with both concussion susceptibility [10,11] and concussion severity [12,13]. *APOE* encodes for the apolipoprotein E (apo-E), a lipid-transporting glycoprotein. The rs405509 TT genotype was over-represented in individuals with a concussion history (odds ratio 2.80, 95% confidence interval: 1.10 – 6.90) [10] and in addition, the *APOE*-ε4 allele was reported to be associated with concussion history in collegiate athletes (odds ratio 9.80, 95% confidence interval: 1.00 – 96.55) [11]. Furthermore, the TT genotype was associated with severe outcomes from concussion (death, vegetative state or severe disability; odds ratio 0.37, 95% confidence interval: 0.13 - 1.01) [12] while *APOE*-ε4 was associated with greater severity for chronic brain injury in boxers ($P=0.040$) [13]. The role of *APOE* rs405509 T and ε4 alleles in increased risk for Alzheimer's disease is well-documented [18,19].

The findings for *APOE* rs405509 single nucleotide polymorphism (SNP) and the *APOE*-ε isoforms, as derived from the rs429358-rs7412 haplotype analyses, have, unfortunately, not been consistently replicated in the various independent studies [20–22]. However, the overall biological evidence from expression studies implicating the *APOE* rs405509 SNP (Ch. 19:44905579; 3105bp from rs429358, 3243bp from rs7412) in protein regulation [14] and *APOE*-ε2/ε3/ε4 isoforms (rs429358: Ch.

19:44908684, rs7412: Ch. 19:44908822; 138bp from each other) in neuron repair [23] still provides support for further investigation into this region.

To date, there have been no published reports investigating all three SNPs (rs405509, rs429358 and rs7412) collectively, and their combined association with concussion risk and severity. Therefore, the aim of this case-control genetic association study was to investigate the association between the *APOE* SNPs (rs405509: G>T, rs429358: T>C, rs7412: C>T) and the ε2/ε3/ε4 isoforms with (i) concussion history and (ii) concussion severity while playing rugby. The *a priori* hypothesis is that the *APOE* rs405509 TT genotype and ε4 allele (i.e. C-C rs429358-rs7412 haplotype) are associated with (i) increased concussion susceptibility and (ii) delayed concussion severity. The novel aspect of this study was to investigate all three variants with concussion history and concussion severity and to evaluate the inferred haplotype of rs405509 and ε2/ε3/ε4 isoforms.

2. Methods

A case-control genetic association study was performed according to the STREGA guidelines for reporting genetic association studies [24]. In total, 696 athletes who played rugby during the 2013 to 2015 rugby seasons were invited to participate from high schools, amateur-level clubs, professional rugby teams and medical practices in South Africa. Ethics approval was obtained from the Health Science Faculty, Human Research Ethics Committee of the University of Cape Town and the Western Cape Education Department. Furthermore, permission to conduct this study was granted by the respective high schools, clubs and university authorities. The study was performed according to the Declaration of Helsinki.

Concussions were defined according to the Zurich Concussion Consensus statement [1]. Briefly, concussed cases were defined as individuals who sustained a concussion from a direct or indirect blow to the head, during a sporting activity. Furthermore, concussions diagnosed by a medical

4

professional with one or more concussion symptoms reported were categorized as a "diagnosed concussion"; whereas, concussions not diagnosed by a medical professional but with one or more symptoms reported were categorized as a "self-reported concussion". Symptoms were selected from the concussion consensus approved list in the Sports Concussion Assessment Tool version 3 (SCAT3), and included 14 physical (e.g. headaches), six cognitive (e.g. loss of consciousness) and four emotional (e.g. irritability) symptoms.

Individuals who (i) signed the informed consent ($\geq$ 18 years old; N=271) or (ii) signed the assent (minor) and consent (parent/guardian) forms (< 18 years old; N=129) and (iii) also completed the information forms (N=400) were included in this study. Participants were excluded from this study based on (i) ancestry, only individuals who self-identified as white were included in this study (N=82 were excluded) because of population stratification effects, (ii) sporting activity (only rugby players were included; N=1 was excluded), and (iii) sex, only males were included (N=314) while three females were excluded. Furthermore, only individuals who sustained a rugby-related concussion were included (N=15 were excluded). Participants self-reporting meningitis (N=4), stroke (N=2) and seizures (N=3) were excluded. Finally, the known relatives of participants (N=2), were excluded. After all exclusions, a total of 288 white, male rugby players (aged 12 – 39 years old) were analysed; with 128 participants self-reporting no history of concussions (control group, CON) and 160 participants with previously diagnosed and self-reported concussions (all cases group, ALL; Supplementary Fig. S1). A subgroup, of the all cases group, was comprised only of the previously diagnosed concussed participants and separately analysed (clinically diagnosed subgroup, CLIN; N=139). All participants were analysed collectively and the junior group (N=121, 12 – 18 year old high school rugby players) was also independently analysed, due to the vulnerability of youth athletes to adverse complications from concussion [7–9], however, as a result of the small sample size in juniors, the exploratory analysis of the junior group was performed and included in the supplementary material (Supplementary Tables S3).

Participants completed details of their four most recent previous concussions (if applicable), and their sporting and medical histories (Supplementary Table S4). Participation in sports was categorized into rugby (in isolation), non-rugby collision sports, non-rugby contact sports and non-contact sports. For rugby participation, participants self-reported the total number of years playing rugby. All participants donated a blood or buccal sample for genetic analysis. DNA was obtained and extracted from either a cheek swab as per the manufacturers' protocols [25] or a 5 ml venous blood sample as previously described [26].

The selected *APOE* gene SNPs, rs405509 (-219 G>T) [14], rs429358 (T>C), rs7412 (C>T) [15], had a > 5% minor allele frequency in the white European population (NCBI, https://www.ncbi.nlm.nih.gov/). The C to T substitution at the first codon position, of the rs429358 and rs7412 SNPs, results in a protein change from arginine (CGC) to cysteine (TGC). The rs429358 (T3937C) and rs7412 (C4075T) haplotype was used to infer the *APOE* isoforms namely: C^{3937}-C^{4075} for ε4 allele, T^{3937}-C^{4075} for ε3, C^{3937}-T^{4075} for ε3r and T^{3937}-T^{4075} for ε2 (Supplementary Fig. S2) [27]. The *APOE* polymorphisms were genotyped, at the Division of Exercise Science and Sports Medicine laboratory, using fluorescence-based TaqMan® real time polymerase chain reaction (PCR) assays and the StepOnePlus™ Real Time PCR machine as per the manufacturer's protocol (Applied Biosystems, CA, USA). Briefly, the cycling conditions used were a holding stage at 95°C for 10 min, 55 cycles of 92°C for 15s and 60°C for 1 min. In a total of 10μl, 2x TaqMan® master mix reagent (Applied Biosystems, CA, USA), 0.25μl of target-specific TaqMan® assay (40x) and 1μl of DNA (6 - 77 ng/μl) were used. Three repeat samples of known genotypes and three samples with no DNA were included as controls. An average genotype call rate of 95% (rs405509: 96%, rs429358: 93% and rs7412: 97%) was noted.

Sample size for this study was determined using QUANTO version 1.2.4 (http://biostats.usc.edu/software). Assuming minor allele frequencies from 0.1 to 0.8, the current study sample size of 160 cases is suitable to detect an allelic odds ratio of 2.0 (as per previous studies), at

80% statistical power and 5% significance level. All images and figures were designed using GraphPad Prism statistical software version 5.03 (GraphPad Software, San Diego, CA, USA). Statistical analysis was performed using the R language and environment [28]. The Hardy-Weinberg equilibrium (HWE) probabilities and genotype frequency distribution differences between groups were determined using the genetics and SNPassoc packages [29]. Where appropriate and hypothesis-driven, the *APOE* rs405509 TT genotype was compared to the combined GG and GT genotype frequencies and GG was also compared to the combined TT and GT. Similarly, for the *APOE* (ε2/ε3/ε4) isoform with the ε4 allele compared against the ε3 and ε2 alleles. Allele frequency differences between groups were determined using Fisher's exact test. The haplo.stats package was used to analyse the rs405509, rs429358 and rs7412 SNPs as an inferred haplotype, with the rs429358-rs7412 haplotype representing the *APOE* (ε2/ε3/ε4) isoform. When all participants were combined, all analyses were adjusted for age as a confounding covariate. In addition to the case-control analysis (i.e. concussion history), the genotype, allele and haplotype frequencies were compared between self-reported concussion symptom duration groups (< 1 week, ≥ 1 week). Symptom duration is used in the clinical decision to determine recovery, which in turn is a proxy for concussion severity [16,17]. Therefore, severity was inferred from symptom duration. When all participants were combined, participants self-reported concussions and symptom duration on average 2.5 ± 2.7 years (N=140) after the concussion occurred. A hypothesis-driven approach was adopted with all polymorphisms located on a single gene, of which rs405509 and rs7412 (D'=0.927, r^2=0.278) and rs429358 and rs7412 (D'=0.995, r^2=0.136) are in high linkage disequilibrium, and therefore correcting for multiple testing would be too conservative [30]. Statistical significance was set at $P<0.05$.

3. Results

The concussed case group (N=160) reported an average of 1.9 ± 1.0 previous concussions (45% sustained one concussion, 34.4% sustained two concussions, 8.8% sustained three concussions, 11.9% sustained four concussions). When all participants (N=188) were combined, the CON group (N=128)

was significantly younger than ALL group (N=160) (*P*=0.008; CON: 19.2 ± 3.5 years old; ALL: 20.5 ± 4.4 years old) and CLIN subgroup (N=139) (*P*=0.015; CLIN: 20.4 ± 4.2 years old). Height, mass, body mass index (BMI), rugby exposure and non-rugby collision sport exposure were not significantly different between groups (Supplementary Table S1).

No genotype association was observed for the *APOE* rs405509, rs429358 and rs7412 SNPs and the *APOE*-ε isoform (ε2, ε3 and ε4 were identified but ε3r was not) on participant characteristics (Supplementary Table S2).

The rs405509 genotype frequency distribution was not significantly different between the CON group and the ALL group or the CLIN subgroup (Table 1). In contrast to the *a priori* hypothesis, the TT genotype was significantly over-represented in the CON group compared to ALL group with 45% reduced concussion susceptibility (TT vs. GG + GT: age-adjusted *P*=0.043; CON: N=35, 29%, ALL: N=28, 18%, odds ratio 0.55, 95% confidence interval: 0.31 − 0.98). No significant rs429358 and rs7412 genotype frequency differences were observed between groups. The rs405509, rs429358 and rs7412 allele frequency distributions were also not significantly different between groups. All three SNPs (rs405509, rs429358 and rs7412) were in HWE for the control and case groups (*P*>0.05).

The allele frequency distribution of *APOE* (ε2/ε3/ε4) isoform was calculated using the inferred *APOE* rs429358-rs7412 haplotypes (Supplementary Fig. S2). The *APOE*-ε isoform frequencies were not significantly different between the control and case groups (data not shown). Furthermore, the *APOE*-ε isoform distribution was in HWE in the control and case groups (*P*>0.05).

Table 1 The genotype, minor allele frequencies (%) and HWE test for the *APOE* rs405509, rs429358 and rs7412 polymorphisms between the control group (CON), diagnosed and self-reported concussed cases (ALL) and medical/clinically diagnosed concussed cases only (CLIN)

		CON	ALL	*P*-value[a]	CLIN	*P*-value[b]
All participants	N	(123)	(153)		(132)	
	GG	26.8 (33)	29.4 (45)	0.121[c]	30.3 (40)	0.172[c]
rs405509	GT	44.7 (55)	52.3 (80)	(**0.043**[d], 0.704[e])	51.5 (68)	(0.063[d], 0.557[e])
	TT	28.5 (35)	18.3 (28)		18.2 (24)	
	T allele	50.8 (125)	44.4 (136)	0.145[c]	43.9 (116)	0.132[c]
	HWE	0.279	0.516		0.724	
	N	(117)	(150)		(129)	
	TT	70.9 (83)	70.0 (105)	0.950[c]	69.0 (89)	0.988[c]
rs429358	CT	27.4 (32)	28.0 (42)		29.5 (38)	
	CC	1.7 (2)	2.0 (3)		1.6 (2)	
	C allele	15.4 (36)	16.0 (48)	0.905[c]	16.3 (42)	0.916[c]
	HWE	1.000	0.768		0.524	
	N	(126)	(153)		(132)	
	CC	81.7 (103)	82.4 (126)	0.361[c]	82.6 (109)	0.404[c]
rs7412	CT	17.5 (22)	17.6 (27)		17.4 (23)	
	TT	0.8 (1)	0.0 (0)		0.0 (0)	
	T allele	9.5 (24)	8.8 (27)	0.771[c]	8.7 (23)	0.933[c]
	HWE	1.000	0.607		0.597	

Genotype and allele frequencies are represented as percentages with the number of participants with non-missing data (N) shown in parentheses. Statistical significant differences (*P*<0.05) between control and case groups are highlighted. [a]CON group vs. ALL group. [b]CON group vs. CLIN subgroup. [c]Age-adjusted global *P*-values (likelihood test, Fisher's exact test) comparing genotype and allelic (minor vs. major: rs405509, T vs. G; rs429358, C vs. T; rs7412, T vs. C) frequencies between groups. Post-hoc: *P*-values for [d]TT vs. GG + GT and [e]GG vs. TT + GT (control vs. case groups), are displayed in parentheses. HWE are *P*-values from exact tests performed for the *APOE* polymorphisms. HWE - Hardy-Weinberg Equilibrium
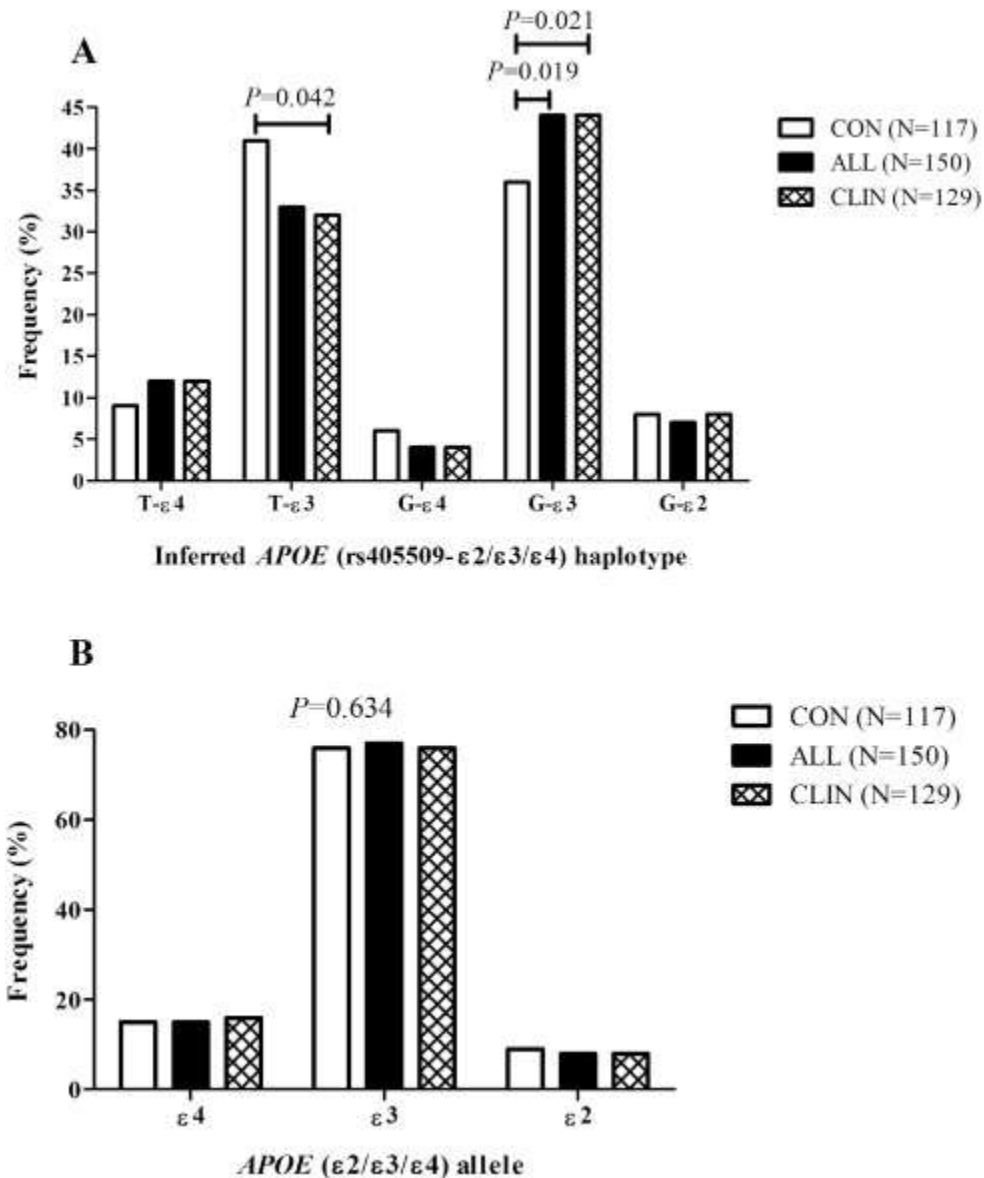
Figure 1 The *APOE* rs405509-ε2/ε3/ε4 inferred haplotype and allele frequency distributions

A) The inferred *APOE* rs405509-ε2/ε3/ε4 haplotype frequency distributions between the non-concussed and concussed participants when all participants were combined. **B)** The *APOE* ε2/ε3/ε4 allele frequency distributions between the non-concussed and concussed participants when all participants were combined. The *APOE*-ε isoforms represent the rs429358 (T>C)-rs7412 (C>T) inferred haplotypes with the ε4 allele representing C-C, ε3 allele T-C and ε2 allele T-T. Self-reported, non-concussed participants (CON), all diagnosed and self-reported concussed cases (ALL) and medical/clinically diagnosed concussed cases (CLIN) are displayed (N=267). Age-adjusted *P*-values are displayed for control vs. case groups (significance set at *P*<0.05).

The inferred *APOE* haplotypes (rs405509-ε2/ε3/ε4) were derived using the genotype data [rs405509 (G>T), rs429358 (T>C), rs7412 (C>T)]. In total, five inferred haplotypes, above a frequency of 4%, were identified (Fig. 1A). The T-ε3 haplotype was significantly over-represented in the CON group (N=48, 41%) compared to CLIN subgroup (N=41, 32%, recessive model: age-adjusted *P*=0.042, hap.score= -2.04). Additionally, the G-ε3 haplotype was significantly under-represented in the CON group (N=42, 36%) compared to the ALL group (N=66, 44%, dominant model: age-adjusted *P*=0.019, hap.score= 2.35) and CLIN subgroup (N=57, 44%, dominant model: age-adjusted *P*=0.021, hap.score= 2.31; Fig. 1A). No significant differences for the *APOE* (ε2/ε3/ε4) allele frequencies were observed between groups (*P*>0.05; Fig. 1B).

Of the total 160 concussed cases, 83% reported on their previous concussion symptom duration (65% reported symptoms persisting < 1 week, 35% reported symptoms persisting ≥ 1 week). Age, height, mass, BMI, rugby exposure and non-rugby collision sport exposure were not significantly different between the concussion symptom duration groups (Supplementary Table S1).

No significant rs405509, rs429358 and rs7412 genotype frequency differences were noted between duration groups (Table 2). In contrast to the *a priori* hypothesis, no significant rs405509 genotype differences between groups were noted. No significant rs405509, rs429358 and rs7412 allele frequency differences were noted, except for the rs405509 in which the T allele was significantly over-represented in the < 1 week duration group (< 1 week: N=70, 50.7%) compared to the ≥ 1 week duration group (≥ 1 week: N=29, 36.3%; odds ratio 0.55, 95% confidence interval: 0.30 – 1.01; *P*=0.048).

Table 2 The genotype frequencies (%) for the *APOE* rs405509, rs429358 and rs7412 polymorphisms between the categories for concussion symptom duration (< 1 week vs. ≥ 1 week)

		Duration of symptoms (ALL)			Duration of symptoms (CLIN)		
		< 1 week	≥ 1 week	*P*-value[a]	< 1 week	≥ 1 week	*P*-value[b]
All participants	N	(83)	(43)		(69)	(40)	
	GG	21.7 (18)	37.2 (16)	0.403	20.3 (14)	40.0 (16)	0.232
rs405509	GT	56.6 (47)	56.6 (21)	(0.584[c], 0.181[d])	58.0 (40)	47.5 (19)	(0.421[c], 0.093[d])
	TT	21.7 (18)	21.7 (6)		21.7 (15)	12.5 (5)	
	T allele	50.0 (83)	38.4 (33)	0.085	50.7 (70)	36.3 (29)	**0.048**
	N	(84)	(42)		(70)	(39)	
	TT	67.9 (57)	69.0 (29)	0.103	67.1 (47)	69.2 (27)	0.128
rs429358	CT	28.6 (24)	31.0 (13)		30.0 (21)	30.8 (12)	
	CC	3.6 (3)	0 (0)		2.9 (2)	0 (0)	
	C allele	17.9 (30)	15.5 (13)	0.724	17.9 (25)	15.4 (12)	0.709
	N	(84)	(43)		(71)	(40)	
rs7412[d]	CC	84.7 (72)	81.4 (35)	0.858	84.5 (60)	82.5 (33)	0.700
	CT	15.3 (13)	18.6 (8)		15.5 (11)	17.5 (7)	
	T allele	7.6 (13)	9.3 (8)	0.637	7.7 (11)	8.8 (7)	0.802

P-value (Fisher's exact tests) comparing the *APOE* rs405509, rs429358 and rs7412 genotype and allele (minor vs. major: rs405509, T vs. G; rs429358, C vs. T; rs7412, T vs. C) frequency distributions between the categories for concussion symptom duration ([a]diagnosed and self-reported concussed case group, ALL and [b]clinically diagnosed concussed case subgroup, CLIN), with significance highlighted in bold (*P*<0.05, adjusted for age and number of previous concussion/s). Post-hoc: [c]TT vs. GG + GT and [d]GG vs. TT + GT, < 1 week vs. ≥ 1 week case groups, are displayed in parentheses. [e]All individuals with the rare rs7412 TT genotype are missing symptom duration data or did not sustain a concussion.

The *APOE*-ε isoform frequency distribution was not significantly different between duration groups (data not shown). In total, four inferred *APOE* haplotypes (rs405509-ε2/ε3/ε4), above a frequency of 4%, were identified for the duration groups. No significant differences were noted for the four inferred *APOE* haplotypes (rs405509-ε2/ε3/ε4) between the duration groups (data not shown).

4. Discussion

The aim of this exploratory study was to investigate the association between the *APOE* SNPs (rs405509, rs429358, and rs7412), concussion history and severity in rugby players. The main preliminary findings of the study noted for all participants collectively, included: (i) the rs405509 TT genotype was significantly associated with a 45% reduced risk of concussion susceptibility (odds ratio 0.55, 95% confidence interval: 0.31 – 0.98), (ii) the novel association of the inferred *APOE* (rs405509-ε2/ε3/ε4) T-ε3 haplotype with reduced concussion risk while the G-ε3 haplotype was associated with increased concussion risk, (iii) no independent association was noted between *APOE* ε2/ε3/ε4 genotypes and concussion susceptibility, and (iv) the rs405509 TT genotype was significantly associated with reduced severity (< one week recovery) following a concussion. These preliminary findings collectively suggest that the *APOE* rs405509 T allele may be associated with reduced concussion susceptibility and the alternate G allele with an increased susceptibility.

The promoter rs405509 (-219 G>T) SNP is located in the transcriptional regulatory region of *APOE* and has been implicated in the regulation of *APOE* expression [14]. In particular, the T allele of rs405509 was shown to stimulate increased promoter activity and nuclear protein binding [14], thereby increasing apo-E protein expression. The TT genotype has been linked with greater amyloid β protein deposition which is associated with Alzheimer's disease pathology [31]. The odds ratio of 0.55 for TT genotype and reduced risk in this study is a similar effect size as previous studies with odds ratios varying from 0.37 to 2.80 [10,12]. The previous studies [11,20,21] which did not replicate the rs405509 and concussion risk association were constrained by small sample sizes and inappropriate methods of statistical analysis for determining genotype distribution. The phenomenon of population stratification, which is accounted for in this study, highlights the importance of only comparing individuals of similar ancestry and may also explain the discrepancies in the findings of previous studies which did not account for this. Although the evidence implicating *APOE* rs405509 locus with

the risk of concussion is conflicting, the evidence from the functional studies [14] still implicate this *APOE* promoter region in neurodegenerative diseases.

For all participants, the T allele was independently associated with rapid recovery (Table 2). Although it was interesting to note that the TT genotype was also associated with rapid recovery while GG genotype was associated with prolonged recovery in juniors, the wide confidence intervals and the small sample size constrained the confidence in these associations in juniors (Supplementary Table S3).

The *APOE*-ε4 allele plays a role in the pathological deposition of amyloid β protein [32]. A previous study has reported an association between the ε4 allele and increased concussion risk [11]. The extremely wide confidence interval and close to non-significant *P*-value (odds ratio 9.8, 95% confidence interval: 1.00 – 96.55, *P*=0.050), limits confidence in the significant association reported [11]. Furthermore, previous studies found no association between the ε4 allele and concussion risk [20,21]. This study was unable to demonstrate an association between ε4 allelic polymorphism with concussion risk. The lack of associations in the current study and previous studies [20,21] could possibly be due to an insufficient sample size to detect the rare ε4 'risk' allele. It is evident that apo-E plays an important biological role within the brain and specifically its isoforms differentially modulate neurite growth, altering the cytoskeleton thereby playing a role in neuronal repair and maintenance [23].

A seminal study implicated the ε4 allele in traumatic brain injury severity in boxers [13]. Conversely, no association was found between the ε4 allele and severity of mild traumatic brain injury in paediatrics [22]. Our current study did not find an association between ε4 allele and prolonged symptom duration in concussed individuals, a proxy for concussion severity. Previous studies which found significance for rs405509 and *APOE*-ε4 with concussion severity investigated traumatic and more severe head injuries compared to the concussions sustained in our cohort, in which no severe disability or debilitating outcome was reported.

This study's genotype and allele frequency distributions were in HWE (Table 1). The rs405509 (T=51%), rs429358 (C=15%) and rs7412 (T=10%) minor allele frequencies, within the CON group, are in alignment with NCBI frequencies for the white population (rs405509 T=49%, rs429358 C=15% and rs7412 T=8%, https://www.ncbi.nlm.nih.gov/projects/SNP/). This implies that the significant genotype associations found in this study are possibly not affected by evolutionary elements (e.g. genetic drift) and population stratification but rather the concussion phenotype.

The T-ε3 and G-ε3 haplotypes were associated with concussion history thereby implicating the T allele in decreased susceptibility and the G allele in increased susceptibility. The T allele increases *APOE* promoter activity, increasing apo-E protein expression which could decrease amyloid β deposition, thereby decreasing or preventing the release of cytokines and reactive oxygen species and preventing neuronal apoptosis. Alternatively, the G allele decreases *APOE* promoter activity, decreasing apo-E protein expression which increases amyloid β deposition thereby eliciting the release of cytokines and reactive oxygen species leading to neuronal apoptosis. We propose that the T allele provides a neuroprotective response to a head knock protecting the brain against the development of concussion signs and symptoms. Additionally, this neuroprotective response could promote a favourable outcome and a rapid recovery following concussion. Whereas, the G allele promotes a neurodegenerative response following a head knock, provoking the concussion signs and symptoms. This neurodegenerative response could lead to a poor outcome and a prolonged recovery following concussion.

Although the results were preliminary, it was interesting to note an independent association between rs405509 T allele and reduced concussion susceptibility was observed when only juniors were evaluated. The small sample size for juniors limits interpretation of this finding. The vulnerability of youth athletes to prolonged recovery and debilitating post-concussion complications [7–9] highlight the need for future studies to evaluate *APOE* genetic profiles and concussion aetiology in a larger cohort of youth athletes.

One of the limitations of this study is the inclusion of self-reported concussions and concussion symptom duration (an average of 2.5 ± 2.7 years recall after the concussion for all participants; 1.0 ± 1.3 years recall after the concussion for juniors), which can be influenced by potentially inaccurate participant recall. This can result in misreporting of concussion and symptom resolution. However, 87% (N=139) of the cases from all participants were diagnosed by a medical professional. Participants self-reporting ancestry could potentially result in inaccurate ancestry identification which multidimensional scaling tools (e.g. Plink) may have more accurately assigned. A reduced statistical power indicates an increased likelihood that the significant associations might be false positives (Type I error). This may be true for the symptom duration association, in juniors, with only 30% power detection (Supplementary Table S3).

5. Conclusion

The observed associations for this *APOE* region and concussion aetiology could be a reflection of linkage disequilibrium with causal polymorphisms located elsewhere on *APOE* or closely-related genes. Therefore, future studies should explore this region or related regions to identify sequence motifs which may be directly related to concussion susceptibility. As well as, further studies in larger cohorts should be performed particularly for the association of *APOE-ε* isoform. The novel finding of this study was the inferred *APOE* rs405509-ε2/ε3/ε4 haplotypes modulating the risk of concussion which provides evidence to support the further elucidation of the *APOE* gene in concussion aetiology.

6. Practical Implications

- *APOE* rs405509 TT genotype may be associated with reduced concussion susceptibility.

- *APOE-ε4* allele was not associated with concussion susceptibility.

- The TT genotype may influence symptom duration following a concussion.

Acknowledgements

References

1.   McCrory P, Meeuwisse WH, Aubry M et al. Consensus Statement on Concussion in Sport-The 4th International Conference on Concussion in Sport Held in Zurich, November 2012. *Br J Sports Med* 2013; 5(4):255–279.

2.   Guilmette TJ, Rasile D. Sensitivity, specificity, and diagnostic accuracy of three verbal memory measures in the assessment of mild brain injury. *Neuropsychology* 1995; 9(3):338–344.

3.   Shaw NA. The neurophysiology of concussion. *Prog Neurol Surg* 2002; 67:281–344.

4.   Gardner AJ, Iverson GL, Williams WH et al. A systematic review and meta-analysis of concussion in rugby union. *Sport Med* 2014; 44(12):1717–1731.

5.   Abrahams S, Mc Fie S, Patricios J et al. Risk factors for sports concussion: an evidence-based systematic review. *Br J Sports Med* 2014; 48(2):91–97.

6.   Makdissi M, Davis G, Jordan B et al. Revisiting the modifiers: how should the evaluation and management of acute concussions differ in specific groups? *Br J Sports Med* 2013; 47(5):314–320.

7.   Hartlage L, Durant-Wilson D, Patch P. Persistent neurobehavioral problems following mild traumatic brain injury. *Arch Clin Neuropsychol* 2001; 16(6):561–570.

8.   Field M, Collins MW, Lovell MR et al. Does age play a role in recovery from sports-related concussion? A comparison of high school and collegiate athletes. *J Pediatr* 2003; 142(5):546–553.

9.   Wetjen NM, Pichelmann MA, Atkinson JLD. Second impact syndrome: Concussion and second injury brain complications. *J Am Coll Surg* 2010; 211(4):553–557.

10.   Terrell TR, Bostick RM, Abramson R et al. *APOE*, *APOE* promoter, and Tau genotypes and risk for concussion in college athletes. *Clin J Sport Med* 2008; 18(1):10–17.

11.   Tierney RT, Mansell JL, Higgins M et al. Apolipoprotein E genotype and concussion in college athletes. *Clin J Sport Med* 2010; 20(6):464–468.

12.   Lendon CL, Harris JM, Pritchard AL et al. Genetic variation of the *APOE* promoter and

outcome after head injury. *Neurology* 2003; 61(5):683–685.

13. Jordan BD. Apolipoprotein E ε4 Associated With Chronic Traumatic Brain Injury in Boxing. *J Am Med Assoc* 1997; 278(2):136–140.

14. Artiga MJ, Bullido MJ, Sastre I et al. Allelic polymorphisms in the transcriptional regulatory region of apolipoprotein E gene. *FEBS Lett* 1998; 421(2):105–108.

15. Zannis VI, Just PW, Breslow JL. Human apolipoprotein E isoprotein subclasses are genetically determined. *Am J Hum Genet* 1981; 33(1):11–24.

16. Erlanger D, Kaushik T, Cantu R et al. Symptom-based assessment of the severity of a concussion. *J Neurosurg* 2003; 98(3):477–484.

17. Makdissi M, Darby DG, Maruff P et al. Natural history of concussion in sport: markers of severity and implications for management. *Am J Sport Med* 2010; 38(3):464–471.

18. Xin X-Y, Ding J-Q, Chen S-D. Apolipoprotein E promoter polymorphisms and risk of Alzheimer's disease: evidence from meta-analysis. *J Alzheimers Dis* 2010; 19(4):1283–1294.

19. Namboori PKK, Vineeth K V., Rohith V et al. The *APOE* gene of Alzheimer's disease (AD). *Funct Integr Genomics* 2011; 11(4):519–522.

20. Terrell TR, Bostick RM, Rogers GL et al. Association of *APOE* and other genetic polymorphisms with prospective concussion risk in a prospective cohort study of college athletes. *Biomed. Sci. Eng. Cent. Conf. (BSEC), 2014 Annu. Oak Ridge Natl. Lab.* Oak Ridge, IEEE, 2014. p. 1–4.

21. Kristman VL, Tator CH, Kreiger N et al. Does the apolipoprotein epsilon 4 allele predispose varsity athletes to concussion? A prospective cohort study. *Clin J Sport Med* 2008; 18(4):322–328.

22. Moran LM, Taylor HG, Ganesalingam K et al. Apolipoprotein E4 as a predictor of outcomes in pediatric mild traumatic brain injury. *J Neurotrauma* 2009; 26(9):1489–1495.

23. Mahley RW, Rall SC. Apolipoprotein E: Far More Than a Lipid Transport Protein. *Annu Rev Genomics Hum Genet* 2000; 1(1):507–537.

24. Little J, Higgins JPT, Ioannidis JPA et al. STrengthening the REporting of genetic association studies (STREGA)- An extension of the STROBE statement. *Genet Epidemiol* 2009;

33(7):581–598.

25. McMichael GL, Gibson CS, O'Callaghan ME et al. DNA from buccal swabs suitable for high-throughput SNP multiplex analysis. *J Biomol Tech* 2009; 20(5):232–235.

26. Mokone GG, Hons M, Gajjar M et al. The Guanine-Thymine Dinucleotide Repeat Polymorphism Within the Tenascin-C Gene Is Associated With Achilles Tendon Injuries. *Sport Med* 2005; 33(7):1016–1021.

27. Seripa D, Panza F, Franceschi M et al. Non-apolipoprotein E and apolipoprotein E genetics of sporadic Alzheimer's disease. *Ageing Res Rev* 2009; 8(3):214–236.

28. R Core Team. *R: A language and environment for statistical computing*. Vienna, Austria., 2013.

29. González JR, Armengol L, Solé X et al. SNPassoc: An R package to perform whole genome association studies. *Bioinformatics* 2007; 23(5):644–645.

30. Nyholt DR. A simple correction for multiple testing for single-nucleotide polymorphisms in linkage disequilibrium with each other. *Am J Hum Genet* 2004; 74(4):765–769.

31. Lambert JC, Mann D, Richard F et al. Is there a relation between *APOE* expression and brain amyloid load in Alzheimer's disease? *J Neurol Neurosurg Psychiatry* 2005; 76(7):928–933.

32. Xu Q, Bernardo A, Walker D et al. Profile and Regulation of Apolipoprotein E (ApoE) Expression in the CNS in Mice with Targeting of Green Fluorescent Protein Gene to the *APOE* Locus. *J Neurosci* 2006; 26(19):4985–4994.

Supplementary material

Table S1: The general characteristics for all participants between the control group (CON), diagnosed and self-reported concussed cases (ALL) and medical/clinically diagnosed concussed cases (CLIN)

	CON	ALL	*P*-value[a]	CLIN	*P*-value[b]	< 1 week	≥ 1 week	*P*-value[a]	< 1 week	≥ 1 week	*P*-value[b]
All participants											
N	(128)	(160)		(139)		(86)	(47)		(72)	(44)	
Age (yrs.)	19.2 ± 3.5 (128)	20.5 ± 4.4 (160)	**0.008**	20.4 ± 4.2 (139)	**0.015**	20.8 ± 4.2 (86)	19.3 ± 3.9 (47)	**0.049**	20.9 ± 4.2 (72)	19.4 ± 3.9 (44)	0.070
Height (m)	1.82 ± 0.08 (126)	1.84 ± 0.08 (158)	0.201	1.83 ± 0.08 (137)	0.308	1.84 ± 0.08 (86)	1.84 ± 0.08 (46)	0.985	1.84 ± 0.09 (72)	1.83 ± 0.08 (43)	0.694
Mass (kg)	89.5 ± 15.0 (125)	92.5 ± 15.9 (156)	0.110	92.5 ± 16.8 (135)	0.122	93.7 ± 16.7 (84)	91.1 ± 15.4 (47)	0.376	94.2 ± 17.6 (70)	90.7 ± 15.8 (44)	0.282
BMI (kg/m²)	26.9 ± 3.5 (124)	27.3 ± 3.5 (155)	0.297	27.4 ± 3.6 (134)	0.262	27.5 ± 3.7 (84)	27.1 ± 3.4 (46)	0.469	27.6 ± 3.8 (70)	27.1 ± 3.5 (43)	0.475
Rugby exposure (yrs.)[c]	11.2 ± 4.1 (128)	11.3 ± 4.4 (159)	0.810	11.2 ± 4.6 (138)	0.996	12.0 ± 4.6 (85)	10.6 ± 4.3 (47)	0.075	11.9 ± 4.6 (71)	10.5 ± 4.5 (44)	0.102
Non-rugby collision sport exposure (yrs.)[d]	5.4 ± 4.5 (7)	5.9 ± 3.4 (11)	0.799	5.9 ± 3.9 (8)	0.840	4.3 ± 1.7 (4)	5.0 ± 3.4 (4)	0.705	3.5 ± 2.1 (2)	5.0 ± 3.4 (4)	0.606
Junior											
N	(62)	(59)		(50)		(28)	(23)		(21)	(21)	
Age (yrs.)	16.6 ± 1.1 (62)	16.5 ± 1.2 (59)	0.621	16.4 ± 1.3 (50)	0.265	16.8 ± 1.3 (28)	16.3 ± 1.3 (23)	0.179	16.6 ± 1.3 (21)	16.3 ± 1.4 (21)	0.486
Height (m)	1.81 ± 0.08 (61)	1.81 ± 0.07 (57)	0.772	1.80 ± 0.08 (48)	0.750	1.80 ± 0.07 (28)	1.82 ± 0.08 (22)	0.625	1.80 ± 0.08 (21)	1.81 ± 0.08 (20)	0.811
Mass (kg)	82.3 ± 14.4 (60)	83.5 ± 15.5 (56)	0.649	82.4 ± 16.5 (47)	0.972	83.9 ± 17.9 (27)	84.2 ± 14.1 (23)	0.936	83.0 ± 19.7 (20)	83.5 ± 14.5 (21)	0.932
BMI (kg/m²)	25.3 ± 3.5 (59)	25.5 ± 3.9 (55)	0.741	25.4 ± 4.0 (46)	0.883	25.5 ± 4.5 (27)	25.8 ± 3.3 (22)	0.806	25.2 ± 4.8 (20)	25.7 ± 3.5 (20)	0.691
Rugby exposure (yrs.)[c]	9.0 ± 2.7 (62)	9.1 ± 3.0 (59)	0.805	8.8 ± 3.4 (50)	0.682	9.6 ± 2.7 (28)	8.6 ± 3.3 (23)	0.237	9.4 ± 3.0 (21)	8.2 ± 3.2 (21)	0.241
Non-rugby collision sport exposure (yrs.)[d]	4.0 ± 2.8 (2)	5.6 ± 2.4 (7)	0.448	5.8 ± 2.8 (5)	0.475	5.0 ± 1.0 (3)	5.7 ± 3.8 (3)	0.783	5.0 (1)	5.7 ± 3.8 (3)	0.893

All values are mean ± standard deviation with the number of participants with non-missing data (N) in parentheses. *P*-values (unadjusted) for the CON group compared to [a]ALL group and [b]CLIN subgroup, with significant differences (*P*<0.05) highlighted. [c]Self-reported number of years playing rugby. [d]Self-reported number of years playing other collision sports excluding rugby e.g. combative sports. BMI - body mass index, yrs. – years

Table S2: The *P*-values for general characteristics of all participants (N=288) between *APOE* rs405509 (G>T), rs429358 (T>C), rs7412 (C>T) and *APOE*-ε (ε2/ε3/ε4) polymorphisms

	rs405509[a]	rs429358[b]	rs7412[c]	*APOE*-ε isoforms[d]
All participants (N)				
Age (yrs.)	0.636	0.111	0.377	0.110
Height (m)	0.622	**0.044** (0.094)[e]	0.255	**0.048** (0.104)[e]
Mass (kg)	0.983	0.074	0.285	0.059
BMI (kg/m²)	0.53	0.253	0.578	0.222
Rugby exposure (yrs.)[f]	0.822	0.099	0.157	0.050
Junior (N)				
Age (yrs.)	0.679	0.510	0.633	0.802
Height (m)	0.838	0.921	0.981	0.938
Mass (kg)	0.968	0.302	0.567	0.659
BMI (kg/m²)	0.805	0.359	0.503	0.758
Rugby exposure (yrs.)[f]	0.994	0.575	0.404	0.385

X^2 test *P*-values (age-adjusted) between [a]rs405509 (T>G), [b]rs429358 (C>T) and [c]rs7412 (T>C) genotypes with significant differences are highlighted (*P*<0.05). [d]*APOE* ε2/ε3/ε4 genotype was determined from the rs429358-rs7412 haplotype. [e]Age-adjusted *P*-value, [f]Self-reported number of years playing rugby. BMI – body mass index, yrs. – years

Table S3: The genotype, minor allele frequencies (%) and HWE test for the *APOE* rs405509, rs429358 and rs7412 polymorphisms between the control group (CON), all medical/clinically diagnosed and self-reported concussed cases (ALL) and medical/clinically diagnosed concussed cases only (CLIN) as well as between the categories for concussion symptom duration (< 1 week vs. ≥ 1 week)

		Concussion history					Duration of symptoms (ALL)			Duration of symptoms (CLIN)		
		CON	ALL	P-value[a]	CLIN	P-value[b]	< 1 week	≥ 1 week	P-value[c]	< 1 week	≥ 1 week	P-value[d]
Junior	N	(57)	(54)		(45)		(25)	(21)		(18)	(19)	
	GG	22.8 (13)	31.5 (17)	**0.036**[d]	33.3 (15)	0.102[d]	20.0 (5)	52.4 (11)	0.085	16.7 (3)	57.9 (11)	**0.005**
rs405509	GT	43.9 (25)	55.6 (30)	(**0.010**[e], 0.303[f])	51.1 (23)	(**0.037**[e], 0.238[f])	64.0 (16)	42.9 (9)	(0.289[c], **0.032**[d])	61.1 (11)	42.1 (8)	(**0.016**[c], **0.008**[d])
	TT	33.3 (19)	13.0 (7)		15.6 (7)		16.0 (4)	4.8 (1)		22.2 (4)	0.0 (0)	
	T allele	55.3 (63)	40.7 (44)	**0.033**[d]	41.1 (37)	**0.049**[d]	48.0 (24)	26.2 (11)	0.052	52.8 (19)	21.1 (8)	**0.007**
	HWE	0.425	0.398		0.425		0.245	1.000		0.635	0.538	
	N	(58)	(55)		(46)		(27)	(21)		(20)	(19)	
	TT	75.9 (44)	83.6 (46)	0.574[d]	84.8 (39)	0.524[d]	81.5 (22)	90.5 (19)	0.744	80.0 (16)	94.7 (18)	0.410
rs429358	CT	20.7 (12)	14.5 (8)		13.0 (6)		14.8 (4)	9.5 (2)		15.0 (3)	5.3 (1)	
	CC	3.4 (2)	1.8 (1)		2.0 (1)		3.7 (1)	0.0 (0)		5.0 (1)	0.0 (0)	
	C allele	13.8 (16)	9.1 (10)	0.302[d]	8.0 (7)	0.282[d]	11.1 (6)	4.8 (2)	0.459	12.5 (5)	2.6 (1)	0.201
	HWE	0.278	0.361		0.278		0.267	1.000		0.246	1.000	
	N	(60)	(55)		(46)		(27)	(21)		(20)	(19)	
	CC	88.3 (53)	78.2 (43)	0.142[d]	78.3 (36)	0.163[d]	81.5 (22)	76.2 (16)	0.706	80.0 (16)	78.9 (15)	0.958
rs7412[h]	CT	11.7 (7)	21.8 (12)		21.7 (10)		18.5 (5)	23.8 (5)		20.0 (4)	21.1 (4)	
	T allele	5.8 (7)	10.9 (12)	0.230[d]	10.9 (10)	0.208[d]	9.3 (5)	11.9 (5)	0.744	10.0 (4)	10.5 (4)	1.000
	HWE	-	-		-		-	-		-	-	

Genotype and allele frequencies are represented as percentages with the number of participants with non-missing data (N) shown in parentheses. Statistical significant differences ($P<0.05$) between control and case groups are highlighted ([a]CON group vs. ALL group. [b]CON group vs. CLIN subgroup), and between the categories for concussion symptom duration < 1 week vs. ≥ 1 week groups ([c]ALL groups and [d]CLIN subgroup, adjusted for age and number of previous concussions). [e]Age-adjusted global *P*-values (likelihood test, Fisher's exact test) comparing genotype and allelic (minor vs. major: rs405509, T vs. G; rs429358, C vs. T; rs7412, T vs. C) frequencies between groups. *A priori* hypothesis (post-hoc): *P*-values for [f]TT vs. GG + GT, [g]GG vs. TT + GT (control vs. case groups, < 1 week vs. ≥ 1 week case groups) are displayed in parentheses. [h]All individuals with the rs7412 TT genotype is missing symptom duration data or did not sustain a concussion. [i]*APOE* ε2/ε3/ε4 genotype was determined from the rs429358-rs7412 haplotype, all individuals with the ε2/ε2 genotype is missing symptom duration data or did not sustain a concussion. HWE are *P*-values from exact tests performed for the *APOE* polymorphisms. HWE - Hardy-Weinberg Equilibrium

Table S4: Proportion (%) of self-reported medical conditions for all participants and between the control group (CON), diagnosed and self-reported concussed cases (ALL)

	Total[a] (N)	CON	ALL
Asthma	32	31.0 (13)	39.6 (19)
Elevated blood cholesterol	22	23.8 (10)	25.0 (12)
Learning disability[b]	16	23.8 (10)	12.5 (6)
Mood disorders	1	2.4 (1)	0.0 (0)
Multiple conditions[c]	8	4.8 (2)	12.5 (6)
Osteoarthritis	2	2.4 (1)	2.1 (1)
Recurring headache/migraine	9	11.9 (5)	8.3 (4)

[a]Total number (N) of participants self-reporting a medical condition. Medical conditions for CON and ALL are displayed as percentages with the number of participants with non-missing data (N) in parentheses. [b]A broad term for self-reporting a learning disability, low learning level, attention deficit disorders or medication prescribed for learning disorders (e.g. Rittelin). [c]Self-reporting more than one medical condition

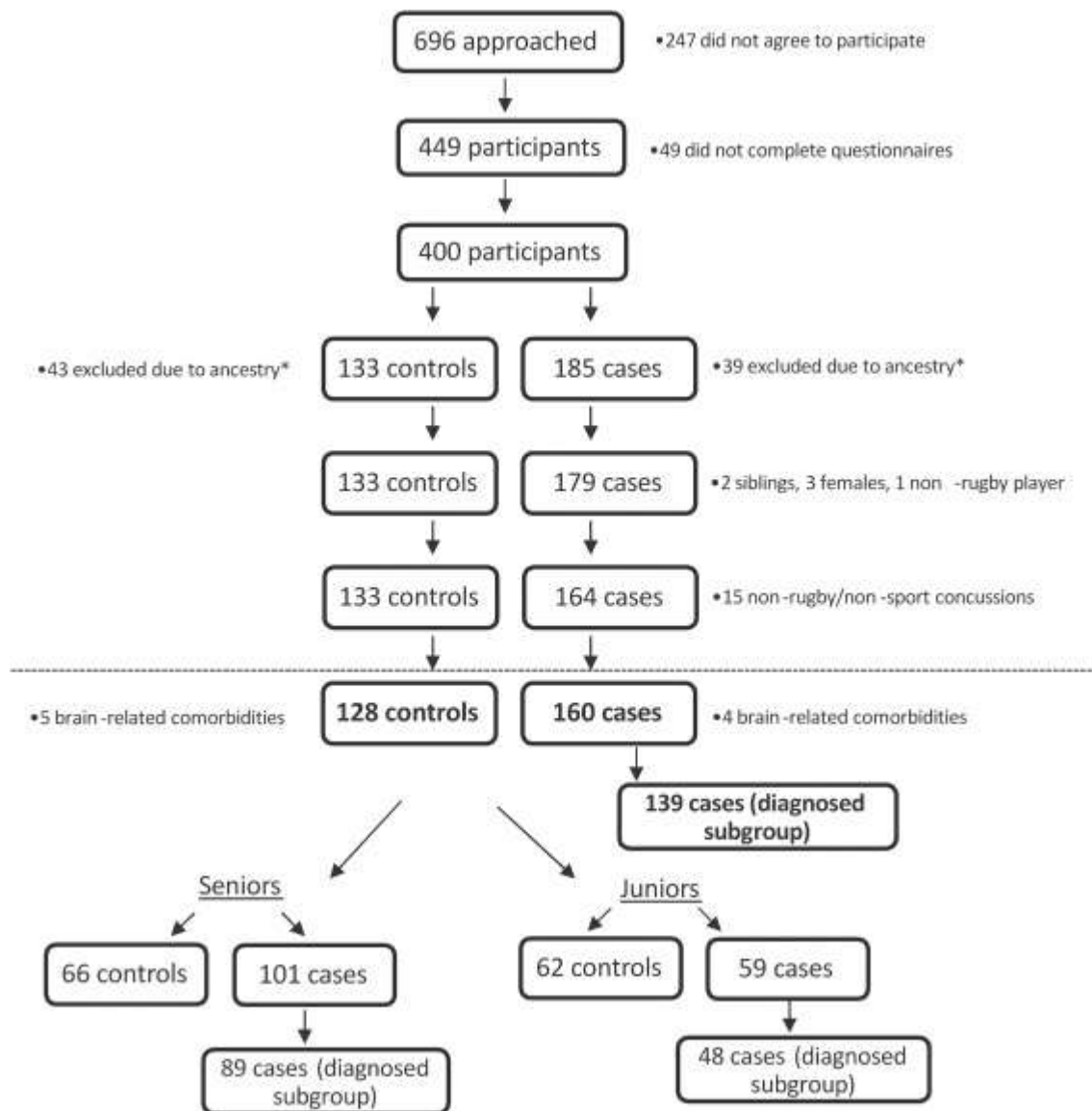**Supplementary figure legends**

Figure S1 Participant recruitment and eligibility

The number of participants included in the study (highlighted in boxes), are displayed as control and case groups, with the number (and reason) of excluded participants (not highlighted). The final total included and analyzed participants are grouped below the dotted lines with 128 control participants and 160 concussed case participants, with a subgroup of 139 diagnosed case participants. *Participants of similar ancestry can only be compared because of population stratification and therefore only participants who had self-reported white ancestry, who presented the majority of the cohort, were analyzed in this study.
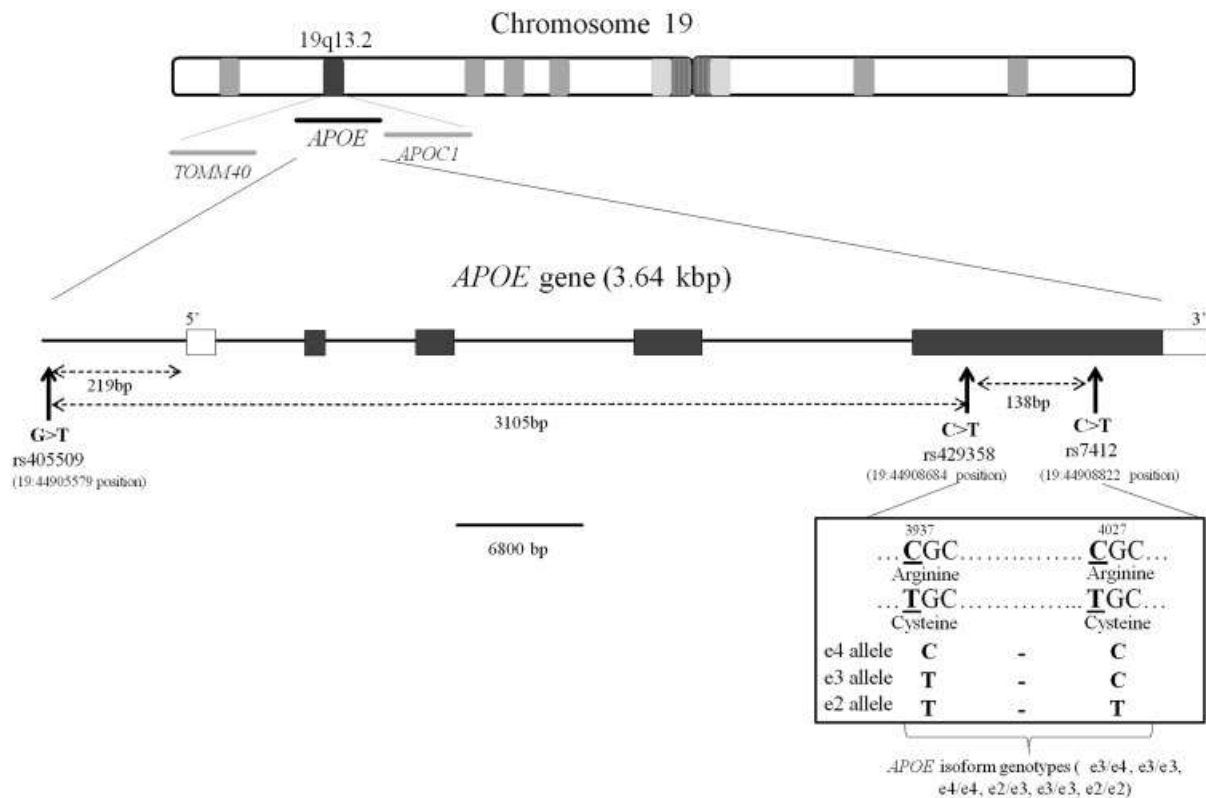
Figure S2 The *APOE* gene

The *APOE* gene is located at chromosome 19q3.2 consisting of four exons (black boxes), with the 5'-and 3'-untranslated regions (UTRs; white boxes) and a one cm scale bar displayed. The rs405509 (G>T), rs429358 (T>C) and rs7412 (C>T) polymorphisms are indicated on the gene with the single nucleotide change, the NCBI accession number, the chromosomal position in parentheses, the distance between polymorphisms and between rs405509 and the promoter region. The single nucleotide and amino acid changes for the rs429358 and rs7412 polymorphisms with the haplotype combinations and the related *APOE*-ε isoforms are displayed (black outlined box). *APOE* – apolipoprotein E gene, *APOC1* – apolipoprotein C1 gene, *TOMM40* - translocase of outer mitochondrial membrane 40 homolog gene, kbp - kilobase pairs, bp - base pairs. This figure was constructed from information obtained using the database hosted by NCBI (http://www.ncbi.nlm.nih.gov/, GRCh38 assembly).