A Genome-wide Association Study for Concussion Risk

STUART K. KIM[1], MEGAN D. ROCHE[2], MICHAEL FREDERICSON[2], JASON L. DRAGOO[3], BRANDON H. HORTON[4], ANDY L. AVINS[4], HEATHER G. BELANGER[5,6,7], JOHN P. A. IOANNIDIS[8,9], and GEOFFREY D. ABRAMS[2]

[1]Department of Developmental Biology, Stanford University Medical School, Stanford, CA; [2]Department Orthopaedic Surgery, Stanford University Medical Center, Stanford, CA; [3]UCHealth Steadman Hawkins Clinic Denver–Surgery Center, Englewood, CO; [4]Division of Research, Kaiser Permanente Northern California, Oakland, CA; [5]United States Special Operations Command (USSOCOM), Tampa, FL; [6]St. Michael's Inc., Tampa, FL; [7]Department of Psychiatry and Behavioral Neurosciences and Department of Psychology, University of South Florida, Tampa, FL; [8]Department of Medicine, Stanford Prevention Research Center, Department of Epidemiology and Population Health, and Department of Biomedical Data Science, Stanford University School of Medicine, Stanford, CA; and [9]Department of Statistics, Stanford University School of Humanities and Sciences, Stanford, CA

ABSTRACT

KIM, S. K., M. D. ROCHE, M. FREDERICSON, J. L. DRAGOO, B. H. HORTON, A. L. AVINS, H. G. BELANGER, J. P. A. IOANNIDIS, and G. D. ABRAMS. A Genome-wide Association Study for Concussion Risk. *Med. Sci. Sports Exerc.*, Vol. 53, No. 4, pp. 704–711, 2021. **Purpose:** This study aimed to screen the entire genome for genetic markers associated with risk for concussion. **Methods:** A genome-wide association analyses was performed using data from the Kaiser Permanente Research Bank and the UK Biobank. Concussion cases were identified based on electronic health records from the Kaiser Permanente Research Bank and the UK Biobank from individuals of European ancestry. Genome-wide association analyses from both cohorts were tested for concussion using a logistic regression model adjusting for sex, height, weight, and race/ethnicity using allele counts for single nucleotide polymorphisms. Previously identified genes within the literature were also tested for association with concussion. **Results:** There were a total of 4064 cases of concussion and 291,472 controls within the databases, with two single nucleotide polymorphisms demonstrating a genome-wide significant association with concussion. The first polymorphism, rs144663795 ($P = 9.7 \times 10^{-11}$; OR = 2.91 per allele copy), is located within the intron of *SPATA5*. Strong, deleterious mutations in *SPATA5* cause intellectual disability, hearing loss, and vision loss. The second polymorphism, rs117985931 ($P = 3.97 \times 10^{-9}$; OR = 3.59 per allele copy), is located within *PLXNA4. PLXNA4* plays a key role is axon outgrowth during neural development, and DNA variants in *PLXNA4* are associated with risk for Alzheimer's disease. Previous investigations have identified five candidate genes that may be associated with concussion, but none showed a significant association in the current model ($P < 0.05$). **Conclusion:** Two genetic markers were identified as potential risk factors for concussion and deserve further validation and investigation of molecular mechanisms. **Key Words:** TRAUMATIC BRAIN INJURY, GENETIC TESTING, SPATA5, PLXNA4

Concussion is a traumatic brain injury induced by biomechanical forces (1). It can result from a direct blow to the head or neck area or from an impacted force to another portion of the body with resultant forces to the head. Concussion is common among athletes involved in collision sports and soldiers, but it can also occur as a result of falls, car accidents, or other events leading to head trauma (2,3). In the United States, up to 3.8 million sports-related concussions are diagnosed yearly (4), whereas in England and Wales, 1.4 million people per year seek care at hospitals after a head injury (5). Between 2000 and 2015, there were more than 250,000 military personnel diagnosed with concussion (6).

Sports-related concussion manifests itself with a variety of overlapping signs and symptoms such as physical maladies, cognitive impairment, neurobehavioral features, and sleep disturbance. With appropriate treatment of first time concussions, approximately 80%–90% of persons experience symptom resolution within 7–10 d (7). Others, particularly those with prior concussions or underlying psychological conditions, can experience prolonged symptoms of months or longer (8). Although recovery is expected after a single, uncomplicated concussion,

Address for correspondence: Stuart K. Kim, Ph.D., Department of Developmental Biology, Stanford University Medical School, Stanford, CA 94305-5329; E-mail: stuartkm@stanford.edu.
Submitted for publication July 2020.
Accepted for publication September 2020.
Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's Web site (www.acsm-msse.org).

0195-9131/20/5304-0704/0
MEDICINE & SCIENCE IN SPORTS & EXERCISE®

DOI: 10.1249/MSS.0000000000002529

there are many factors that can influence recovery, such as age (9), sex (10–12), history of prior concussions (12,13), education level (14), premorbid mental health disorders (11,12,15), and comorbid mood disorder (16). Repetitive concussions have been linked with long-term neurodegenerative issues, including chronic traumatic encephalopathy (17,18).

Previous studies have reported genetic risk factors that are associated with an increased risk of concussion (19–22). These studies have taken a candidate gene approach using genes with known neurological functions, with many focusing on the E4 allele of the apolipoprotein E (ApoE) gene (19–25). Other candidate genes that have been tested for association with concussion risk include genes involved in neuronal function (DRD4, SLC17A7, BDNF, and NGF), neuronal disease (MAPT), and neuroinflammation (IL-6/IL-6R) (19,23,25–29).

In contrast to candidate gene studies that are restricted to testing a limited number of preselected genes, genome-wide association (GWA) analyses reveal the strongest associations in the genome by searching in an agnostic fashion without prior bias. In a GWA analysis, millions of polymorphisms throughout the genome are scanned for potential associations with a phenotype such as concussion, in which a variant allele is more prevalent in cases than in controls. To achieve statistical significance, extremely large cohorts of cases and controls are required. Uncovering genetic markers for risk for concussion could help in understanding the biological mechanisms underlying concussion as well as symptoms associated with postconcussion syndrome. Further, genetic markers could be used to identify individuals at risk. This information might lead to improved prevention and treatment for concussion, help inform return-to-play guidelines, and perhaps direct an athlete's choice of sport or training.

In this work, we report the first GWA analysis for concussion using data from the Kaiser Permanente Research Bank (KPRB) and the UK Biobank. We hypothesize that genes without previously documented association with concussion would be identified.

METHODS

The genome-wide association study (GWAS) for concussion was performed using data from the KPRB (with which the Kaiser Permanente, Northern California Research Program on Genes, Environment, and Health [RPGEH] is affiliated) and from the v3 release of UK Biobank.

KPRB cohort. The data generation and data analysis pipelines for the KPRB cohort have been previously described (30). Our analysis cohort includes 83,414 individuals of European ancestry who were genotyped at 670,572 single nucleotide polymorphisms (SNP) using Affymetrix Axiom genome-wide arrays. Genotypes were prephased with Shape-IT v2.r644 (accessed February 2, 2016) then imputed to a cosmopolitan reference panel consisting of all individuals from the 1000 Genomes Project (Mar 2012 release) using IMPUTE2 v2.2.2 (accessed February 2, 2016) and standard procedures with a cutoff of $R^2 > 0.3$. The final number of SNP after imputation was 9,878,560. The quality of the imputed data was previously validated (31).

UK Biobank cohort. Genotype data were obtained from the v3 release of UK Biobank (32). The UK Biobank electronic health care records were available for 212,122 individuals and included data until June 2019. Genotype data were imputed centrally by UK Biobank with IMPUTE2 using the Haplotype Reference Consortium and the UK10k + 1000GP3 reference panels (33). Metrics for quality control were established and then used to filter DNA variants by UK Biobank (32). Imputed SNP were excluded if they had an IMPUTE2 info score <0.4.

Database quality control. For both the KPRB and the UK Biobank cohorts, individuals were excluded if they were outliers based on genotyping missingness rate or heterogeneity, whose sex inferred from the genotypes did not match their self-reported sex, who withdrew from participation, or who were not of European ancestry. The purpose of restricting individuals to those with European ancestry is to reduce population stratification in the study; for example, if the risk of concussion among individuals with African ancestry is higher than that for European individuals, then any SNP with an allele frequency that is different between African and European ancestries would appear to be associated with concussion. Overall, these filters resulted in excluding 18.9% and 7.9% of individuals (mostly due to the ancestry filter) in the KPRB and UK Biobank cohorts, respectively. Genetic variants that failed quality control procedures in any of the genotyping batches, that showed a departure from Hardy–Weinberg of $P < 10^{-50}$, or that had a minor allele frequency < 0.001 were excluded. The determination of genetic ancestry was performed by principal component analysis computed centrally by either KPRB or UK Biobank, as previously described (32).

Phenotype definitions. In the KPRB cohort, concussion cases were identified based on clinical diagnoses captured in the Kaiser Permanente Northern California electronic health record system before July 22, 2015. The *International Classification of Diseases, Ninth Revision*, or the *International Classification of Diseases, Tenth Revision*, codes were used to identify cases of concussion (Table 1). In the UK Biobank cohort, concussion cases were identified from primary care data, using either Read v2 or Read v3 codes (Table 1).

GWA. GWAS was conducted using PLINK v2.0a (30). SNP associations were tested with concussion with a logistic regression model using allele counts for typed and imputed SNP. The model was adjusted for genetic sex, height, weight, and race/ethnicity using 10 principal components, as previously described (30). For the UK Biobank, the age of injury was available and was included as an adjustment as well, with age of enrollment used for controls. The final number of SNP that was analyzed was 9,878,560 in the KPRB cohort and 9,856,942 in the UK Biobank cohort. To account for inflation due to population stratification, the genomic control parameter (λ_{gc}) was calculated ($\lambda_{gc} = 1.056$ for KPRB; $\lambda_{gc} = 0.925$ for UK Biobank). Subsequently, P values were adjusted for the genomic control in each population.

TABLE 1. Phenotype codes.

	Code Description	Number
KPRB		
ICD-9 code		
850	Concussion without loss of consciousness	513
850.11	Concussion, with loss of consciousness of 30 min or less	373
850.5	Concussion with loss of consciousness	379
850.9	Concussion, unspecified	1227
854.09	Intracranial injury of other and unspecified nature without mention of open intracranial wound, with concussion, unspecified	108
310.2	Traumatic (postconcussion) encephalopathy	1058
ICD-10 code		
F07.81	Postconcussion syndrome	319
S06.0X0A	Concussion without loss of consciousness, initial encounter	286
S06.0X0D	Concussion without loss of consciousness, subsequent encounter	19
S06.0X1A	Concussion with loss of consciousness of 30 min or less, initial encounter	90
S06.0X1D	Concussion with loss of consciousness of 30 min or less, subsequent encounter	18
S06.0X9A	Concussion with loss of consciousness of unspecified duration, initial encounter	430
S06.0X9D	Concussion with loss of consciousness of unspecified duration, subsequent encounter	139
S06.0X9S	Concussion with loss of consciousness of unspecified duration, sequela	17
S14.0XXA	Concussion and edema of cervical spinal cord, initial encounter	7
	Total unique cases KPRB	3170
	Total controls KPRB	80,244
UK Biobank		
read2		
S60	Concussion	185
S600	Concussion with no loss of consciousness	49
S601	Concussion with less than 1 h loss of consciousness	1
S602	Concussion with 1–24 h loss of consciousness	0
S603	Concussion with more than 24 h loss of consciousness and return to preexisting conscious level	0
S604	Concussion with more than 24 h loss of consciousness without return to preexisting conscious level	0
S605	Concussion with loss of consciousness of unspecified duration	1
S60z	Concussion NOS	2
e2a2	Postconcussion syndrome	65
read3		
S60	Concussion	409
S600	Concussion with no loss of consciousness	178
S601	Concussion with less than 1 h loss of consciousness	11
S602	Concussion with 1–24 h loss of consciousness	2
S603	Concussion with more than 24 h loss of consciousness and return to preexisting conscious level	0
S604	Concussion with more than 24 h loss of consciousness without return to preexisting conscious level	1
S605	Concussion with loss of consciousness of unspecified duration	3
S60z	Concussion NOS	41
e2a2	Postconcussion syndrome	25
XE1bs	Postconcussion syndrome	4
XE2b7	Postconcussion syndrome	141
	Total cases UKB	894
	Total controls UKB	211,228

ICD-9, *International Classification of Diseases, Ninth Revision*; ICD-10, *International Classification of Diseases, Tenth Revision*.

Results using odds ratios per allele from each cohort were combined by inverse-variance, fixed-effects meta-analysis as previously described (30). Here, meta-analysis refers to a statistical method to combine data from GWAS performed on two independent cohorts. The total number of SNP from either the KPRB or the UK Biobank analysis was 14,721,556; the final number of SNP present in both cohorts, and hence analyzed in the fixed-effects meta-analysis, was 7,743,986. For a P value threshold, we set a threshold such that the family-wise error rate is 0.05 using the algorithm developed by Kemp et al. (34); using the total number of SNP from either cohort (14,721,556 SNP) yields $P < 7.7 \times 10^{-9}$ as a threshold for genome-wide significance. Power calculations were conducted with the software using the Genetic Association Study Power Calculator (accessed January 11, 2020) (35).

Further bioinformatics investigation of the top genome-wide significant loci from the GWAS was conducted. QQ and Manhattan plots were created using qqman. Regional association plots were generated for each locus with LocusZoom (accessed January 1, 2020) (36). Neither rs144663795 nor rs117985931 were directly genotyped on the Affymetrix chips, but rather their genotype data were imputed. The info scores from IMPUTE2 were 0.92 and 0.89 for rs144663795 and rs117985931, respectively. These info scores indicate that the imputed genotypes are accurate.

The statistical associations for KPRB and UK Biobank were compared to determine if the association with concussion was stronger in one cohort than the other, a phenomenon referred to as heterogeneity (37). rs144663795 did not show strong evidence for heterogeneity between the KPRB and the UK Biobank cohorts. For rs117985931, Q and I^2 were 0.0 and 0.89, suggesting that the effects were larger in the UK Biobank than the KPRB cohort.

The genomic context of each SNP was investigated using RegulomeDB (accessed February 1, 2020) (38) Web tools. Whether each SNP is an expression quantitative trait locus (eQTL) was queried using the NCBI eQTL Browser (accessed February 1, 2020) and the Genotype-Tissue Expression (GTEx) Portal (accessed February 1, 2020). ChIP seq data from the ENCODE project were used to determine whether SNP were located within transcription factor binding sites (39). Summary statistics for all SNP from the GWAS will be available at NIH GRASP (https://grasp.nhlbi.nih.gov/FullResults.aspx) upon acceptance of this manuscript.

Testing of previously identified candidate genes. A literature search on the genetics of concussion was performed in December 2019 using "concussion genetics" as search terms in PubMed. In addition, references from two recent review article on concussion genetics were included (22,40). Candidate genes were tested for validation when the relevant polymorphisms were present in the summary statistics from the genome-wide analyses performed using the KPRB and the UK Biobank data (Table 4).

Ethical approval. This study analyzed stored data from KPRB and UK Biobank subjects who consented to genomic testing and use of their genomic data, as well as health data from the KPNC and UK Biobank electronic health records. The health and genotype data for the subjects were deidentified. All study procedures were approved by the Institutional Review Board of the Kaiser Foundation Research Institute.

Data sharing. All data will be openly and publicly available upon publication of this article.

Patient involvement. Patients were not involved in this research, except as anonymized subjects in the two cohorts.

RESULTS

Identification of DNA variants associated with concussion. We performed GWA analyses for concussion with the KPRB (83,414 individuals) and UK Biobank (212,122 individuals) cohorts using sex, weight, height, and age of injury as adjustments (Table 2). For KPRB, there were 3170 cases of concussion and 80,244 controls (Table 1). For UK Biobank, there were 894 cases and 211,228 controls (Table 1). Manhattan and QQ plots are shown in Supplemental Figure 1 (see Figure, Supplemental Digital Content 1, QQ and Manhattan plots for genome-wide association analysis of concussion, http://links.lww.com/MSS/C166). There were two SNP with genome-significant associations with concussion using $P = 7.7 \times 10^{-9}$ as a cutoff: rs144663795 on chromosome 4 and rs117985931 on chromosome 7 (Table 3, Fig. 1A and B). rs117985931 shows an association with concussion with a P value ($P = 3.97 \times 10^{-9}$) that is close to the threshold for genome-wide significance, and thus one should be cautious regarding this association until it can be validated in follow-up studies.

Validation of previous candidate gene studies. The E2, E4, and rs405509 alleles of ApoE have been tested for association with risk of concussion in several previous studies, with mixed results (11–13,15–17,41–44). In our analysis, there were a total of 4064 cases of concussion, and neither ApoE4 allele ($P = 0.19$) nor rs405509 ($P = 0.31$) was significantly associated with concussion (Table 4). For ApoE2, the association with concussion has a P value that is borderline significant ($P = 0.067$) with an odds ratio of 1.11 (95% confidence interval = 0.99–1.53).

Previous candidate gene studies have also reported associations with concussion for BDNF, DRD4, SLC17A7, and NGF (19,26). We tested the polymorphisms in these genes in our concussion data set, and we did not find a significant association for any ($P > 0.05$) (Table 4). In summary, we were not able to replicate the results based on candidate genes from any of the previous studies of concussion.

DISCUSSION

Genetic markers for concussion. This study provides new information about a possible genetic mechanism associated with concussion risk. We demonstrated the first evidence for genetic factors affecting risk for concussion with genome-wide significance. Two SNP were associated with concussion in the GWAS (rs144663795 and rs117985931). Previously identified candidate genes within the literature were not confirmed to be associated with concussion in the current data set.

rs144663795 is in the intron of the spermatogenesis associated 5 gene (*SPATA5*), which encodes an ATPase with unknown function. Strong, deleterious missense mutations in *SPATA5* cause severe intellectual disability, hearing loss, vision impairment, and slow and asymmetric waves in the EEG (45–47). *SPATA5* function may be weakly affected by rs144663795 with no overt phenotype except for an increased risk of concussion, and pathologic mutations in *SPATA5* may directly cause a strong neurological phenotype.

rs117985931 is located in an intron of *PLXNA4*, which encodes a plexin that is a type of neuronal receptor whose role is to transduce signals from semaphorins to steer axon growth (48). Neural injury causes an increase in the expression of *PlexinA4*, suggesting a role in neural maintenance and regeneration (41). A GWAS revealed that DNA variants in *PLXNA4* are associated with Alzheimer's disease (42,43). The finding from this study adds an important new insight into the overall roles for plexins, linking their roles in neural development, neural injury, Alzheimer's disease, and concussion.

Individuals harboring risk alleles for rs144663795 (C) in *SPATA5* or rs117985931 (G) in *PLXNA4* have an increased risk for concussion. The frequencies of rs144663795 (C) in *SPATA5* and rs117985931 (G) were 1.4% and 0.13% in the UK Biobank cohort, respectively (Table 3). Although the genetic association results have not yet been validated in an independent study, the two genetic markers could provide key information to athletes and soldiers about their risk for concussion. Genetic testing in either athletes or soldiers could allow them to take extra precautions to avoid concussion and also to compel them to seek clinical treatment that they might otherwise ignore. The genetic information could also be used by medical professionals to make more informed decisions regarding concussion diagnosis, management, and return to

TABLE 2. Study demographics.

	Case	Control	*P*	Risk (%)
KPRB				
Male	1652	41,879		3.8
Female	2536	57,523	0.0006*	4.2
Height (inches)	65.65 ± 9.43	65.73 ± 10.17	0.6745	
Weight (lb)	168 ± 37	172 ± 39	<1E–4	
UK Biobank				
Male	479	95,992		0.50
Female	416	115,612	<1E–4*	0.36
Height (inches)	66.6 ± 6.6	66.3 ± 3.6	0.0067	
Weight (lb)	176 ± 37	172 ± 35	0.0009	
Age of injury	40.2 ± 20.6	56.7 ± 8.0	<1E–4	

Data are presented as mean ± SD.
* *P* value vs male.

TABLE 3. Concussion GWAS and meta-analysis.

SNP	Chr	BP	Gene	EA	MAF_KPRB	MAF_UKB
rs144663795	4	124072484	*SPATA5*	C	0.0022	0.014
rs117985931	7	131852272	PLXNA4	G	0.0013	0.0013
KPRB						
SNP	OR_RPGEH (95% CI)	*P*_KPRB				
rs144663795	2.65 (1.86 to 3.78)	1.39E–07				
rs117985931	1.71 (0.99 to 2.92)	5.79E–02				
UK Biobank						
SNP	OR_UKB (95% CI)	*P*_UKB				
rs144663795	4.77 (2.11 to 10.74)	3.21E–05				
rs117985931	11.55 (5.8 to 22.9)	1.10E–14				
Meta-analysis						
SNP	OR_meta	*P*_fixed	*Q*	I^2 (95% CI)		
rs144663795	2.91	9.74E–11	0.46	0 (0 to 0.90)		
rs117985931	3.59	3.97E–09	0	0.89 (0.15 to 0.99)		

CI, confidence interval; Chr, chromosome; BP, base pair; Coordinates use Genome Reference Consortium Human Build 37; EA, effect allele; MAF, minor allele frequency; OR, odds ratio; *Q*, Cochran's *Q*.

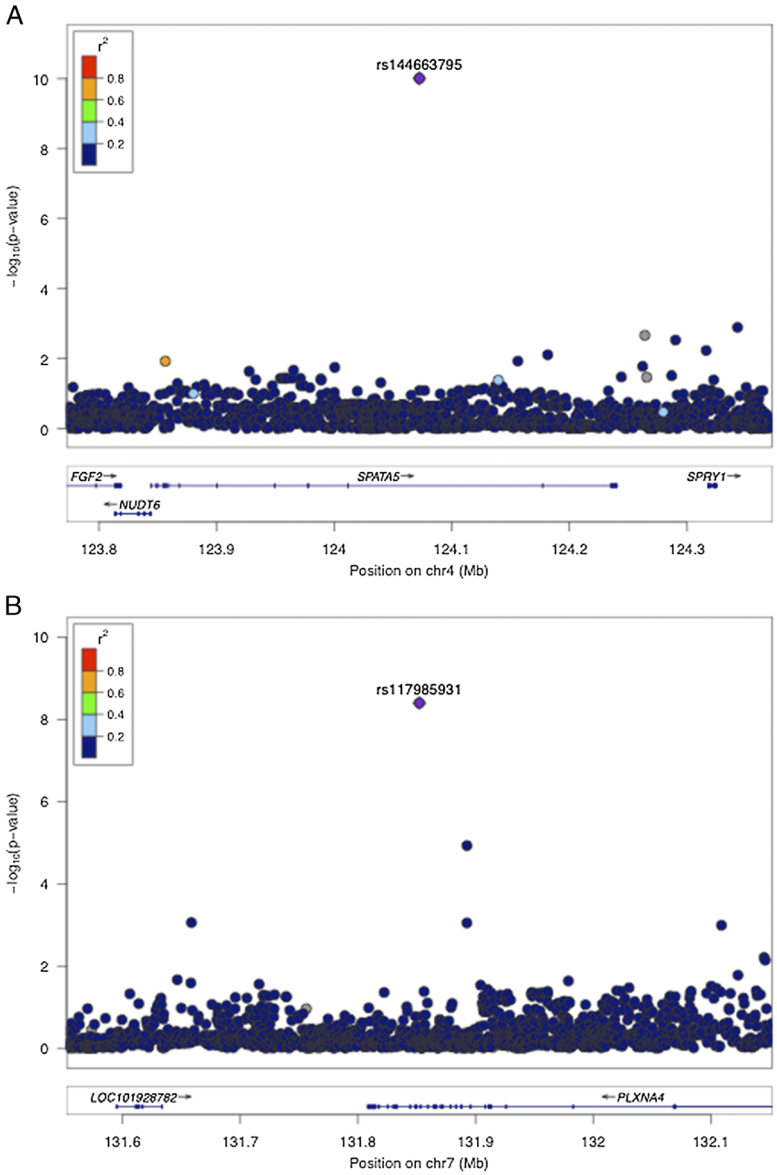

FIGURE 1—A, Regional association plot for rs144663795 with concussion. Tested SNP are arranged by genomic position on chromosome 4 (*x*-axis) in a 600-kb window around the lead SNP rs144663795 (*purple diamond*). The *y*-axis indicates $-\log_{10}$ *P* values for association with concussion for each SNP. rs144663795 is located in the intron of *SPATA5*. The color of dots of the flanking SNP indicates their linkage disequilibrium (R^2) with the lead SNP as indicated by the heat map color key. B, Regional association plot for rs117985931 with concussion. Tested SNP are arranged by genomic position on chromosome 7 (*x*-axis) around the lead SNP rs117985931 (*purple diamond*). The *y*-axis indicates $-\log_{10}$ *P* values for association with concussion for each SNP. rs117985931 is located in an intron of the *PLXNA4* gene. The color of the dots of the flanking SNP indicates their linkage disequilibrium (R^2) with the lead SNP as indicated by the heat map color key.

play. Hercher et al. (44) studied whether NCAA athletes would be interested in genetic testing for concussion and found that 74% expressed interest in undergoing genetic testing for an increased risk of poor recovery from concussion.

Validation of candidate gene studies. Previous studies have tested candidate genes for association with concussion, based on the known role of those genes with neurological disorders such as Alzheimer's disease, major depressive disorder, and impulsive personality traits. In a study involving college athletes, ApoE4 was found to have a significant association (P = 0.04) with the E4 allele acting to protect against concussion (OR = 0.61) (49). In another study involving soldiers, the E4 allele had the opposite effect; it was associated with an increased risk of concussion that was borderline significant (P = 0.087) (19). Other studies found that the E4 allele was associated with either a worse chronic brain injury score (P = 0.04) (20) or a more severe total concussion symptom score (P = 0.015) (21). Finally, two studies found no significant association of the E4 allele with concussion (23,24). In contrast to ApoE4, the ApoE2 allele is associated with decreased risk of Alzheimer's disease (50,51). A third allele of ApoE (rs405509) is referred to as the G-219T promoter polymorphism. One study reported a significant association of rs405509 with concussion (P = 0.04) (52), but two other studies with larger cohorts found no significant association (23,25).

Previous candidate gene studies have reported associations with concussion for BDNF (rs6265) and the CC genotype of

TABLE 4. Validation of previous genetic associations with concussion.

		Previous Work					This Work		
Gene	Allele	Cases	*P*	OR	Phenotype	Ref	Cases	*P**	OR (95% CI)
ApoE	E2 v E3						3515	0.067	1.12 (0.99–1.53)
ApoE	E3 v E4	58	0.99	ND	Concussion history	15	3515	0.19	0.90 (0.76–1.06)
ApoE	E3 v E4	165	0.087	1.5	Concussion history	11	3515	0.19	0.90 (0.76–1.06)
ApoE	E3 v E4	133	0.04	0.61	Concussion history	17	3515	0.19	0.90 (0.76–1.06)
ApoE	E3 v E4	28	0.68	ND	Concussion history	16	3515	0.19	0.90 (0.76–1.06)
ApoE	E3 v E4	45	0.015	5.25	Total symptom score	13	3515	0.19	0.90 (0.76–1.06)
ApoE	E3 v E4	30	0.04	ND	Chronic brain injury score	12	3515	0.19	0.90 (0.76–1.06)
ApoE	rs405509	58	0.16	ND	Concussion history	15	3515	0.31	1.02 (0.87–1.08)
ApoE	rs405510	133	0.31	ND	Concussion history	17	3515	0.31	1.02 (0.87–1.08)
ApoE	rs405511	48	0.04	8.4	Concussion history	44	3515	0.31	1.02 (0.87–1.08)
ApoE	rs405509	97	0.03	2.7	Concussion history	17	1718	0.30	0.95 (0.86–1.04)
BDNF	r6265	165	0.049	2.48	Concussion history	11	3515	0.14	0.95 (0.80–1.11)
DRD4	rs1800955	163	0.014	ND	Concussion history	18	1148	0.99	1.00 (0.84–0.119)
SLC17A7	rs74174284		0.018	ND	Concussion duration and severity	20	3515	0.81	0.99 (0.83–1.15)
NGF	rs3783988	196	0.01	ND	Better functional outcomes	19	3515	0.63	1.02 (0.87–1.17)

DRD4 (rs1800955) (19,26). Other studies have reported an association between the CC genotype of SLC17A7 (rs74174284) and concussion duration and severity and the TT genotype of NGF (rs3783988) with better functional outcomes after concussion (27,28).

Our GWAS contains many more cases of concussions than were available in previous studies (Table 4). For the candidate genes reported in previous studies, we performed follow-up experiments using the GWAS results in an attempt to validate the results using a larger cohort. We were not able to replicate any of the previous reports of association of candidate genes with concussion. Power calculations indicate an 80% chance of success if the genotype relative risk of the candidate gene is at least 1.15. One explanation for the lack of validation is that the previous studies looked at cases of concussion in athletes and soldiers, whereas our study looked at individuals from the general population. Nevertheless, evidence from many other studies suggests that candidate gene associations need to be independently replicated; otherwise, their credibility is low (53,54).

We found that sex, height, and weight had opposite effects on risk of concussion in the RPGEH and UK Biobank cohorts. In the RPGEH cohort, being female and weighing less were associated with increased risk of concussion, consistent with previous epidemiological studies showing that female athletes are at greater risk for concussion than their male counterparts (55). In the KPRB cohort, however, we found the opposite result; being male, heavier in weight, and taller in height resulted in a higher chance of having had a concussion. The association of weight with concussion may be related to the prior observation that obesity in athletes results in increased time to return to baseline after sports-related concussions (56).

There was a higher incidence of concussion cases in the KPRB cohort (3.8%) than in the UK Biobank cohort (0.42%). The electronic records for both cohorts extend for the entire lifetime of the patient if reported by the patient and recorded by the physician. It is unclear whether the difference in concussion frequency reflects a true difference in incidence or whether it is a bias in how concussion is diagnosed in the Bay Area versus the United Kingdom.

Limitations. Our analysis found only two genome-wide significant signals, possibly because concussion may be poorly documented in these cohorts. This type of misclassification error would mostly tend to dilute the strength of any signals, if present. Alternatively, it could be that the heritability of concussion phenotypes is low. Another limitation is that the phenotypes were defined from codes contained in electronic health records, and thus we have no information regarding the clinical scenarios surrounding the event. This would include whether patients had prior concussions that were not captured and had a series of subconcussive events that went unreported and the force and/or impact velocity of the inciting event. Furthermore, without this clinical information and given the potential inaccuracies of the diagnosis codes within the database, we may be unable to determine whether our identified genetic risk factors place people at increased risk of an initial concussion (i.e., lower their threshold for concussion) or rather place them at increased risk of sustaining more severe and prolonged symptoms (i.e., postconcussive syndrome). In addition, the cohort included people regardless of whether or not they participated in a sport. For example, we were unable to discriminate if the concussions identified in this study were related to participation in sports or from other causes, such as falls, being struck by or against an object, motor vehicle accidents, or self-harm. Lastly, this study only evaluated individuals from the European ancestry group, and the effect of rs144663795 in *SPATA5* or rs117985931 in *PLXNA4* in other ethnicities is unknown.

CONCLUSION

In the future, it will be important to replicate these gene association results with concussion in independent cohorts, especially for athletes and soldiers. Additional studies are warranted to begin to illuminate the underlying biological mechanism for the association of variation near *SPATA5* or *PLXNA4* with concussion. The results from these studies may validate whether rs144663795 in *SPATA5* or rs117985931 in *PLXNA4* can be used as diagnostic markers to help predict which athletes harbor a higher risk for incidence of concussion or from complications after concussion. Follow-up experiments could look at whether rs144663795 in *SPATA5* or rs117985931 in *PLXNA4* affect other aspects of concussion, such as length of recovery time, response to different types of treatment, or risk for chronic traumatic encephalopathy.

This work was supported by grants from the NIH (5RO1AG025941 and RC2 AG036607), the Robert Wood Johnson Foundation, the Wayne and Gladys Valley Foundation, the Ellison Medical Foundation, and the Kaiser Permanente Community Benefits Program.

The authors are deeply indebted to the UK Biobank for providing access to a rich data source and to the access team for assistance with using the data (Application 17847; "GWAS for risk for sports injuries"). They thank Erik Ingelsson for sharing the database at Stanford containing UK Biobank genotype data and Chris Chang for help with PLINK. The authors thank the Kaiser Permanente Northern California KPRB team for access to data and assistance in data management. The data used in this study were provided by the KPRB from the KPRB collection, which includes the Kaiser Permanente RPGEH. Access to data used in this study may be obtained by application to the KPRB via kp.org/researchbank/researchers. A subset of the GERA cohort consented for public use can be found at NIH/dbGaP: phs000674.v3.p3.

S. K. K. is the CEO of AxGen, Inc., a genetic testing company for sports injuries. M. D. R. is a consultant for AxGen, Inc. The results of the present study do not constitute endorsement by the American College of Sports Medicine. The results in this work are presented clearly, honestly, and without fabrication, falsification, or inappropriate data manipulation.

REFERENCES

1. McCrory P, Meeuwisse W, Dvorak J, et al. Consensus statement on concussion in sport—the 5th International Conference on Concussion in Sport held in Berlin, October 2016. *Br J Sports Med*. 2017;51(11):838–47.
2. Baldwin GT, Breiding MJ, Dawn Comstock R. Epidemiology of sports concussion in the United States. *Handb Clin Neurol*. 2018;158:63–74.
3. Cosgrave M, Williams S. The epidemiology of concussion in professional rugby union in Ireland. *Phys Ther Sport*. 2019;35:99–105.
4. Langlois JA, Rutland-Brown W, Wald MM. The epidemiology and impact of traumatic brain injury: a brief overview. *J Head Trauma Rehabil*. 2006;21(5):375–8.
5. Davis T, Ings A. Head injury: triage, assessment, investigation and early management of head injury in children, young people and adults (NICE guideline CG 176). *Arch Dis Child Educ Pract Ed*. 2015;100(2):97–100.
6. Helmick KM, Spells CA, Malik SZ, Davies CA, Marion DW, Hinds SR. Traumatic brain injury in the US military: epidemiology and key clinical and research programs. *Brain Imaging Behav*. 2015;9(3):358–66.
7. McCrory P, Johnston K, Meeuwisse W, et al. Summary and agreement statement of the 2nd International Conference on Concussion in Sport, Prague 2004. *Br J Sports Med*. 2005;39(4):196–204.
8. Iverson GL, Gardner AJ, Terry DP, et al. Predictors of clinical recovery from concussion: a systematic review. *Br J Sports Med*. 2017; 51(12):941–8.
9. Jacobs B, Beems T, Stulemeijer M, et al. Outcome prediction in mild traumatic brain injury: age and clinical variables are stronger predictors than CT abnormalities. *J Neurotrauma*. 2010;27(4):655–68.
10. Meares S, Shores EA, Taylor AJ, et al. Mild traumatic brain injury does not predict acute postconcussion syndrome. *J Neurol Neurosurg Psychiatry*. 2008;79(3):300–6.
11. Ponsford J, Cameron P, Fitzgerald M, Grant M, Mikocka-Walus A, Schonberger M. Predictors of postconcussive symptoms 3 months after mild traumatic brain injury. *Neuropsychology*. 2012;26(3):304–13.
12. Ponsford J, Willmott C, Rothwell A, et al. Factors influencing outcome following mild traumatic brain injury in adults. *J Int Neuropsychol Soc*. 2000;6(5):568–79.
13. Belanger HG, Spiegel E, Vanderploeg RD. Neuropsychological performance following a history of multiple self-reported concussions: a meta-analysis. *J Int Neuropsychol Soc*. 2010;16(2):262–7.
14. Dougan BK, Horswill MS, Geffen GM. Athletes' age, sex, and years of education moderate the acute neuropsychological impact of sports-related concussion: a meta-analysis. *J Int Neuropsychol Soc*. 2014;20(1):64–80.
15. Meares S, Shores EA, Taylor AJ, et al. The prospective course of postconcussion syndrome: the role of mild traumatic brain injury. *Neuropsychology*. 2011;25(4):454–65.
16. Vasterling JJ, Brailey K, Proctor SP, Kane R, Heeren T, Franz M. Neuropsychological outcomes of mild traumatic brain injury, post-traumatic stress disorder and depression in Iraq-deployed US Army soldiers. *Br J Psychiatry*. 2012;201(3):186–92.
17. Bieniek KF, Blessing MM, Heckman MG, et al. Association between contact sports participation and chronic traumatic encephalopathy: a retrospective cohort study. *Brain Pathol*. 2020;30(1):63–74.
18. Goldstein LE, Fisher AM, Tagge CA, et al. Chronic traumatic encephalopathy in blast-exposed military veterans and a blast neurotrauma mouse model. *Sci Transl Med*. 2012;4(134):134ra60.
19. Dretsch MN, Silverberg N, Gardner AJ, et al. Genetics and other risk factors for past concussions in active-duty soldiers. *J Neurotrauma*. 2017;34(4):869–75.
20. Jordan BD, Relkin NR, Ravdin LD, Jacobs AR, Bennett A, Gandy S. Apolipoprotein E epsilon4 associated with chronic traumatic brain injury in boxing. *JAMA*. 1997;278(2):136–40.
21. Merritt VC, Arnett PA. Apolipoprotein E (APOE) ϵ4 allele is associated with increased symptom reporting following sports concussion. *J Int Neuropsychol Soc*. 2016;22(1):89–94.
22. Panenka WJ, Gardner AJ, Dretsch MN, Crynen GC, Crawford FC, Iverson GL. Systematic review of genetic risk factors for sustaining a mild traumatic brain injury. *J Neurotrauma*. 2017;34(13):2093–9.
23. Cochrane GD, Sundman MH, Hall EE, et al. Genetics influence neurocognitive performance at baseline but not concussion history in collegiate student-athletes. *Clin J Sport Med*. 2018;28(2):125–9.
24. 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):322–8.
25. 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):10–7.
26. Abrahams S, McFie S, Lacerda M, et al. Unravelling the interaction between the DRD2 and DRD4 genes, personality traits and concussion risk. *BMJ Open Sport Exerc Med*. 2019;5(1):e000465.
27. Chuang PY, Conley YP, Poloyac SM, et al. Neuroglobin genetic polymorphisms and their relationship to functional outcomes after traumatic brain injury. *J Neurotrauma*. 2010;27(6):999–1006.
28. Madura SA, McDevitt JK, Tierney RT, et al. Genetic variation in SLC17A7 promoter associated with response to sport-related concussions. *Brain Inj*. 2016;30(7):908–13.
29. Mc Fie S, Abrahams S, Patricios J, Suter J, Posthumus M, September AV. Inflammatory and apoptotic signalling pathways and concussion severity: a genetic association study. *J Sports Sci*. 2018;36(19):2226–34.
30. Kim SK, Roos TR, Roos AK, et al. Genome-wide association screens for Achilles tendon and ACL tears and tendinopathy. *PLoS One*. 2017;12(3):e0170422.
31. Jorgenson E, Makki N, Shen L, et al. A genome-wide association study identifies four novel susceptibility loci underlying inguinal hernia. *Nat Commun*. 2015;6:10130.
32. Bycroft C, Freeman C, Petkova D, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature*. 2018;562(7726):203–9.
33. Howie B, Marchini J, Stephens M. Genotype imputation with thousands of genomes. *G3 (Bethesda)*. 2011;1(6):457–70.
34. Kemp JP, Morris JA, Medina-Gomez C, et al. Identification of 153 new loci associated with heel bone mineral density and functional involvement of GPC6 in osteoporosis. *Nat Genet*. 2017;49(10):1468–75.
35. Skol AD, Scott LJ, Abecasis GR, Boehnke M. Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nat Genet*. 2006;38(2):209–13.

36. Pruim RJ, Welch RP, Sanna S, et al. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics*. 2010; 26(18):2336–7.
37. Ioannidis JP, Patsopoulos NA, Evangelou E. Heterogeneity in meta-analyses of genome-wide association investigations. *PLoS One*. 2007;2(9):e841.
38. Boyle AP, Hong EL, Hariharan M, et al. Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res*. 2012;22(9):1790–7.
39. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012;489(7414):57–74.
40. McKeithan L, Hibshman N, Yengo-Kahn AM, Solomon GS, Zuckerman SL. Sport-related concussion: evaluation, treatment, and future directions. *Medical sciences*. 2019;7(3):44.
41. Spinelli ED, McPhail LT, Oschipok LW, Teh J, Tetzlaff W. Class A plexin expression in axotomized rubrospinal and facial motoneurons. *Neuroscience*. 2007;144(4):1266–77.
42. Han Q, Sun YA, Zong Y, et al. Common variants in PLXNA4 and correlation to CSF-related phenotypes in Alzheimer's disease. *Front Neurosci*. 2018;12:946.
43. Jun G, Asai H, Zeldich E, et al. PLXNA4 is associated with Alzheimer disease and modulates tau phosphorylation. *Ann Neurol*. 2014;76(3):379–92.
44. Hercher LS, Caudle M, Griffin J, Herzog M, Matviychuk D, Tidwell J. Student-athletes' views on APOE genotyping for increased risk of poor recovery after a traumatic brain injury. *J Genet Couns*. 2016; 25(6):1267–75.
45. Kurata H, Terashima H, Nakashima M, et al. Characterization of SPATA5-related encephalopathy in early childhood. *Clin Genet*. 2016;90(5):437–44.
46. Tanaka AJ, Cho MT, Millan F, et al. Mutations in SPATA5 are associated with microcephaly, intellectual disability, seizures, and hearing loss. *Am J Hum Genet*. 2015;97(3):457–64.
47. Buchert R, Nesbitt AI, Tawamie H, et al. SPATA5 mutations cause a distinct autosomal recessive phenotype of intellectual disability, hypotonia and hearing loss. *Orphanet J Rare Dis*. 2016;11(1):130.
48. Janssen BJ, Malinauskas T, Weir GA, Cader MZ, Siebold C, Jones EY. Neuropilins lock secreted semaphorins onto plexins in a ternary signaling complex. *Nat Struct Mol Biol*. 2012;19(12):1293–9.
49. Terrell TR, Abramson R, Barth JT, et al. Genetic polymorphisms associated with the risk of concussion in 1056 college athletes: a multicentre prospective cohort study. *Br J Sports Med*. 2018;52(3):192–8.
50. Corder EH, Saunders AM, Risch NJ, et al. Protective effect of apolipoprotein E type 2 allele for late onset Alzheimer disease. *Nat Genet*. 1994;7(2):180–4.
51. Bertram L, McQueen MB, Mullin K, Blacker D, Tanzi RE. Systematic meta-analyses of Alzheimer disease genetic association studies: the AlzGene database. *Nat Genet*. 2007;39(1):17–23.
52. Tierney RT, Mansell JL, Higgins M, et al. Apolipoprotein E genotype and concussion in college athletes. *Clin J Sport Med*. 2010;20(6):464–8.
53. Siontis KC, Patsopoulos NA, Ioannidis JP. Replication of past candidate loci for common diseases and phenotypes in 100 genome-wide association studies. *Eur J Hum Genet*. 2010;18(7):832–7.
54. Ioannidis JP, Tarone R, McLaughlin JK. The false-positive to false-negative ratio in epidemiologic studies. *Epidemiology*. 2011;22(4):450–6.
55. Dick RW. Is there a gender difference in concussion incidence and outcomes? *Br J Sports Med*. 2009;43(1 Suppl):i46–50.
56. Lee YM, Wu A, Zuckerman SL, et al. Obesity and neurocognitive recovery after sports-related concussion in athletes: a matched cohort study. *Phys Sportsmed*. 2016;44(3):217–22.