Thieme

A Genetic Marker Associated with Shoulder Dislocation

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

Affiliations
1 Dev. Bio., Stanford University, Stanford, United States
2 Department of Orthopaedic Surgery, Stanford University, Stanford, United States
3 Orthopedic Surgery, Stanford University, Stanford, United States
4 Division of Research, Kaiser Permanente Northern California, Oakland, United States
5 Department of Medicine, Stanford University, Stanford, United States

Key word
genome-wide association study, sports genetics, shoulder dislocation, genetic epidemiology

accepted after revision 03.02.2017

Bibliography
DOI http://dx.doi.org/10.1055/s-0043-106190
Published online: 18.5.2017
Int J Sports Med 2017; 38: 508–514

ISSN 0172-4622

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

ABSTRACT

Shoulder dislocations are common shoulder injuries associated with athletic activity in contact sports, such as football, rugby, wrestling, and hockey. Identifying genetic loci associated with shoulder dislocation could shed light on underlying mechanisms for injury and identify predictive genetic markers. To identify DNA polymorphisms associated with shoulder dislocation, a genome-wide association screen was performed using publically available data from the Research Program in Genes, Environment and Health including 662 cases of shoulder dislocation and 82 602 controls from the European ancestry group. rs12913965 showed an association with shoulder dislocation at genome-wide significance ($p = 9.7 \times 10^{-9}$; odds ratio = 1.6) from the European ancestry group. Individuals carrying one copy of the risk allele (T) at rs12913965 showed a 69 % increased risk for shoulder dislocation in our cohort. rs12913965 is located within an intron of the TICRR gene, which encodes TOPBP1 interacting checkpoint and replication regulator involved in the cell cycle. rs12913965 is also associated with changes in expression of the ISG20 gene, which encodes an antiviral nuclease induced by interferons. This genetic marker may one day be used to identify athletes with a higher genetic risk for shoulder dislocation. It will be important to replicate this finding in future studies.

Introduction

The shoulder joint has the greatest range of motion of any joint in the body, which makes it susceptible to dislocation. About half of all shoulder dislocations occur while participating in sports or a sports-related activity, especially contact sports such as football, rugby, wrestling and hockey [1, 11, 19, 35]. Shoulder dislocations occur more often in males than females, occur most frequently between the ages of 15 and 30 when people are most physically active, and occur more frequently in players with shoulder hyperlaxity [35]. Once a shoulder dislocation has occurred, the presence of a greater tubercle fracture is associated with decreased risk for a recurrent injury [35].

Anterior shoulder dislocations are the most common, nearly always caused by a traumatic blow or collision that forces the humeral head anteriorly. Posterior shoulder dislocations are less common, and can be caused by convulsive disorders or epileptic seizures. Inferior shoulder dislocations are rare, and are caused by a hyperabduction of the arm that forces the humeral head against the acromion. Shoulder dislocation, especially in young patients, may lead to instability associated with recurrent dislocation, which can be treated with an arthroscopic procedure known as capsulorrhaphy [4, 33].

Very little is known about genetic differences that affect one's inherent risk for shoulder dislocation. A recent study tested a single polymorphism (rs180012) in col1A1 for association with shoulder dislocation (126 cases) and anterior cruciate ligament rupture (233 cases)[20]. rs180012 did not show a significant association with shoulder dislocation alone, but it showed an association when the shoulder dislocation and anterior cruciate ligament rupture cases were combined [20]. To further explore whether there are

genetic differences that affect one's risk for shoulder dislocation, we performed the first genome-wide association study for shoulder dislocation.

Methods

A genome-wide association screen was performed for shoulder dislocation using data from the genotyped Genetic Epidemiology Research on Adult Health and Aging (GERA) cohort of the Kaiser Permanente, Northern California (KPNC) integrated health system. The data generation and data analysis pipeline have been previously described [29]. A complete description of the cohort and study design can be found in dbGaP (Study Accession: phs000674.v1.p1).

Our analysis cohort (n = 102 979) includes 59 671 females, 43 242 males, and 66 individuals of ambiguous sex. Moreover, our analysis cohort is ethnically diverse, including 83 264 European-White (EUR); 8,560 Latino (LAT); 7 518 East Asian (EAS); 3 161 African-American (AFR); and 476 South Asian (SAS) individuals based on ancestry principle components.

Participants were genotyped at over 650 000 SNPs on 4 race/ethnicity-specific Affymetrix Axiom genome-wide arrays optimized for individuals of European (EUR), African-American (AFR), East Asian (EAS), and Latino (LAT) race/ethnicity [12, 13]. Genotype quality control procedures for the GERA cohort were performed as described previously, except that only SNPs with a minor allele frequency of > 1 % were analyzed [22]. 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 imputed to a cosmopolitan reference panel consisting of all individuals from the 1 000 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] and standard procedures with a cutoff of $R^2 > 0.3$ [15]. The quality of the imputed data was previously validated in Jorgenson et al., 2015 [18]. 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 R^2 metric from IMPUTE2 estimates the correlation between the true and imputed genotype and was used to filter SNPs in our GWA analyses [23].

Phenotype definition

Shoulder dislocation cases were identified in the GERA cohort based on clinical diagnoses and surgical procedures captured in the KPNC Electronic Health Record system (▶ **Table 1**). 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.

3 types of shoulder dislocation were included based on ICD-9 codes: anterior (831.01), posterior (831.02) and inferior (831.01) (▶ **Table 1**). Recurrent dislocations, open dislocations and multidirectional instability were not included. Cases of shoulder dislocation were also identified from patients that had undergone a surgical procedure termed capsulorrhaphy that is used to treat shoulder instability, typically resulting from one or more shoulder dislocations. Capsulorrhaphy is often considered even in first-time dislocations [2]. However, some instances of shoulder instability treated by capsulorrhaphy may not involve a current or prior shoulder dislocation, but these instances are likely to be a minor fraction of the cohort. Patients with shoulder capsulorrhaphy were identified as having one or more of 5 Current Procedural Terminology (CPT-4) codes (▶ **Table 1**).

▶ **Table 1** Shoulder dislocation phenotypes in the EUR ancestry group classified by ICD and/or CPT codes.

Code[a]	Code Description	N
ICD831.01	Closed anterior dislocation of humerus	186
ICD831.02	Closed posterior dislocation of humerus	16
ICD831.03	Closed inferior dislocation of humerus	20
CPT29806	Arthroscopy, shoulder, surgical; capsulorrhaphy	155
CPT23450	Capsulorrhaphy, anterior; Putti-Platt procedure or Magnuson type operation	323
CPT23455	Capsulorrhaphy, anterior; with labral repair (eg, Bankart procedure)	355
CPT23460	Capsulorrhaphy, anterior, any type; with bone block	323
CPT23462	Capsulorrhaphy, anterior, any type; with coracoid process transfer	324

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

Some patients had multiple codes associated with shoulder dislocation (▶ **Table 1**). 10 EUR patients had both an anterior and inferior dislocation (ICD831.01 and ICD831.03). 4 capsulorrhaphy CPT codes (CPT23450, CPT23455, CPT23460, CPT23462) overlapped nearly entirely; specifically, 323 patients in the EUR ancestry group had all 4 codes. Of the 186 patients in the EUR ancestry group with ICD831.01 (closed anterior dislocation of humerus), 4 had a capsulorrhaphy procedure.

Genome-wide association

Power calculations were made using the software at http://csg.sph.umich.edu/abecasis/cats/gas_power_calculator/index.html; accessed Dec. 20, 2016 [32]. Genome-wide association analyses of the GERA cohort genomic data were conducted using PLINK v1.90(b3.34) (https://www.cog-genomics.org/plink2; accessed Feb. 1, 2016) [5, 27]. 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 variations in genotyping protocol include: chip type for all populations (refers to the Affymetrix chip version), genotyp-

▶ **Table 2** Demographic factors used in genome-wide association analyses of shoulder dislocation.

	Cases[a]	Controls	Overall
Subjects (%)	662 (0.80 %)	82 602 (99.20 %)	83 264
Sex (%)[b]			
Male	325 (0.93 %)	34 759 (99.07 %)	35 084
Female	335 (0.70 %)	47 783 (90.30 %)	48 118
Undetermined	0 (0 %)	66 (100 %)	66
Age[c]	62.4 (61.5–63.4)	63.9 (63.8–64.0)	63.9 (63.8–64.0)

[a] Cases with shoulder dislocation from the EUR ancestry group; [b] Sex/gender as determined by an individual's genetic data, reported as the number and percentage of total; [c] Average age at subject enrollment (95 % CI)

ing package (refers to the set of chips that were processed together) and reagent kit (refers to A vs O reagent kit distributed by Affymetrix) when there was variation in the individuals of a specific population. We used 10 principal components for EUR, 6 for LAT and 2 for EAS. To account for inflation due to population stratification, the genomic control parameter (λgc) was calculated. λgc is defined as the median of the resulting chi-squared test statistics divided by the expected median of the chi-squared distribution [8]. λgc = 1.0174, 0.958 and 0.933 for EUR, LAT and EAS, respectively. Subsequently, p-values were adjusted for λgc.

There were 626 cases of shoulder dislocation in the EUR ancestry group, 76 in LAT, 45 in EAS, 15 in AFR and 2 in SAS. We used multiple logistic regression models including 11 to 15 covariates, which means that the AFR and SAS ancestry groups could not be analyzed because the number of covariates equal or exceeded the number of cases. For LAT and EAS, we performed the logistic regression but found that the results were not informative due to the relatively small number of cases. We used the EUR ancestry group as the primary source for our analysis.

For the main analysis, an additive model was used in which AA, Aa and aa genotypes are coded as 0, 1 or 2 in the logistic regression that searches for SNPs where 2 effect alleles have a higher risk for shoulder dislocation than one effect allele. One SNP, rs12913965, showed a genome-wide association with shoulder dislocation using the additive model in EUR. For this SNP, we also tested a dominant and recessive model. Specifically, the AA, Aa and aa genotypes are coded as 1, 1, 0 in the dominant model and 1, 0, 0 in the recessive model. That is, one effect allele gives full risk for shoulder dislocation in the dominant model and 2 effect alleles are required to increase the risk in the recessive model.

To determine whether the association of rs12913965 with shoulder injury was different in males vs. females, we performed the logistic regression on males and females separately from the EUR ancestry group. The analysis for each sex was the same as the combined analysis except that sex was no longer used as a covariate and only 3 principal were used. Summary statistics for all of the SNPs from the logistic regression using the additive model are at NIH GRASP: https://grasp.nhlbi.nih.gov/FullResults.aspx.

For the sake of inclusiveness, we performed 2 fixed-effects meta-analyses using inverse variance weights. We combined data from the EUR and LAT ancestry groups in a 2-way meta-analysis, and then data from the EUR, LAT and EAS ancestry groups in a 3-way meta-analysis.

Further bioinformatics investigation of the top genome-wide significant loci from the logistic regression was conducted. The genomic context of each SNP was investigated using RegulomeDB (http://regulomedb.org/; accessed June 1, 2016)[3] and HaploReg (http://archive.broadinstitute.org/mammals/haploreg/haploreg.php; accessed June 1, 2016)[34] web tools. 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 June 2, 2016) and the Genotype-Tissue Expression (GTEx) Portal (http://www.gtexportal.org/home/; accessed June 4, 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 June 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 Institutional Review Board of the Kaiser Foundation Research Institute. This paper conforms to the ethical standards established by this journal [10].

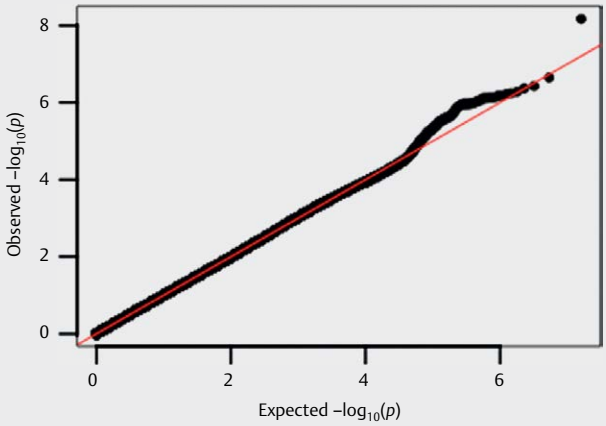
Results

Study population and genotype information

We performed a logistic regression for SNPs associated with shoulder dislocation using genotype and medical data from individuals of the EUR, LAT, AFR, EAS and SAS ancestry groups. The number of cases in the AFR and SAS ancestry groups were too small to analyze, and the number of cases in the LAT and EAS ancestry groups was too small to be informative (Methods). We proceeded to analyze the 83 264 individuals in the EUR ancestry group that included 662 cases and 82 602 controls as our primary analysis. We used an additive model, in which the AA, Aa and aa genothypes are coded as 0, 1 or 2. A description of the sex and age of the cases and controls is shown in ▶ **Table 2**. The International Classification of Diseases, version 9 (ICD-9) as well as Current Procedural Terminology-4 (CPT4) codes used to identify cases from the electronic medical records are shown in ▶ **Table 1**. Overall, the period prevalence of shoulder dislocation was 0.80 %.

We compared the observed p-values from the logistic regression for the EUR ancestry group in a Q-Q plot (▶ **Fig. 1**). We observed deviation from the null hypothesis for the top SNP indicating that this SNP has a lower p-value than would be expected by chance.

The p-value for every SNP from the analysis for the EUR ancestry group is shown in a Manhattan plot in ▶ **Fig. 2**. 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, 25, 26]. One SNP (rs12913965) shows an association with shoulder dislocation that is genome-wide significant using an additive model (▶ **Table 3**). As expected, sex and age of recruitment into the GERA cohort, which are 2 factors known to be associated with shoulder dislocation, showed significant associations as co-variates in the logistic regression for rs12913965 ($p = 3.1 \times 10^{-4}$ and $p = 8.6 \times 10^{-4}$, respectively).

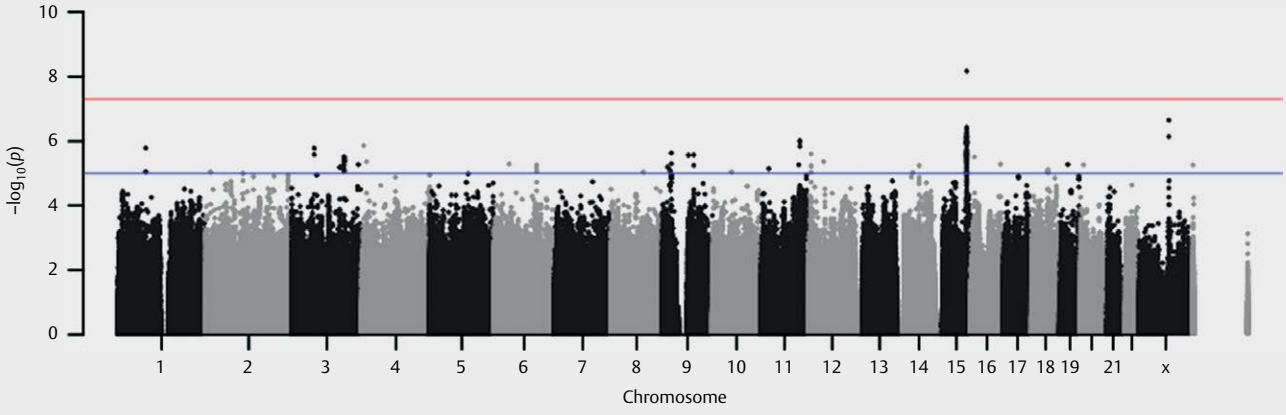


▶ **Fig. 1** Quantile-quantile plot for genome-wide association analyses of shoulder dislocation. The expected vs. observed log transformed values for the 8,046,964 p-values are graphed for association with shoulder dislocation in the EUR ancestry group. The observed p-values (black dots) are plotted on the y-axis and the p-values expected by chance (red line) are plotted on the x-axis.

We tested rs12913965 using either a dominant or recessive model, and found a genome-wide significant association with the dominant but not the recessive model (Methods, ▶ **Table 3**). Our cohort does not let us distinguish between the dominant and additive models. The dominant model predicts that the risk for dislocation for T/T is the same as for T/C at rs12913965, whereas the additive model predicts that the risk for T/T is greater than T/C. However, there were only 10 cases that were T/T at rs12913965, which is too small a number to make a significant distinction between dominant and additive models. rs12913965 was not directly genotyped, but rather the genotype data was imputed with fairly high accuracy ($R^2 = 0.91$)(Supplemental **Table S1**).

We split the EUR ancestry cohort into males and females, and then tested the association of rs12913965 with shoulder dislocation for each sex separately (Supplemental **Table S2**). The odds ratio for females was slightly higher than for males, but the difference was not statistically significant. As would be expected from splitting the cohort in 2, the p-values for males and females are weaker than the p-value for the combined cohort.

For rs12913965, the frequency of the risk allele (T) was 13 % in Europeans and the allelic odds ratio for the additive model was 1.6 (▶ **Table 3**). Individuals that carried one (T) risk allele had a 69 % greater risk of shoulder dislocation in the GERA cohort compared to individuals with no risk alleles (▶ **Table 4**).



▶ **Fig. 2** Manhattan plot for genome-wide association analyses of shoulder dislocation in the EUR ancestry group. The $-\log_{10}$ p-values for association with shoulder dislocation for SNPs from the logistic regression are plotted by genomic position with chromosome number listed across the bottom. The y-axis shows the $-\log_{10}$ p-value for association with shoulder dislocation. The blue line represents suggestive genome-wide significance ($p < 1 \times 10^{-5}$) and the red line represents genome-wide significance ($p < 5 \times 10^{-8}$). The last chromosome is the Y chromosome, which is not labeled.

▶ **Table 3** Genome-wide significant association for shoulder dislocation in the EUR ancestry group.

SNP	Gene	EA[a]	EAF[b]	Test[c]	P-value[d]	OR (95 % CI)[e]
rs12913965	TICCR/ISG20	T	0.10	Add.	9.7×10^{-9}	1.60 (1.37–1.88)
rs12913965	TICCR/ISG20	T	0.10	Dom.	3.6×10^{-9}	1.72 (1.43–2.05)
rs12913965	TICCR/ISG20	T	0.10	Rec.	0.146	1.60 (0.85–3.00)

[a] Effect allele [b] Effect allele frequency in the control population; [c] Additive, Dominant or Recessive model with respect to the effect allele; [d] P-value from fixed-effects logistic regression; [e] Adjusted allelic odds ratio with 95 % confidence interval

rs12913965 is located within an intron of the TICRR gene, which encodes TOPBP1 interacting checkpoint and replication regulator involved in the replication phase of the cell cycle (▶ **Fig. 3**) [31]. Using the Genotype-Tissue Expression database, rs12913965 has not been found to alter expression of the TICRR gene in any of the cell lines tested [7]. However, rs12913965 is associated with changes in expression of the ISG20 gene, which encodes an antiviral nuclease induced by interferons [9]. The ISG20 gene is located about 13 genes and1 Mb to the left of rs12913965 on chromosome 15. The T (risk) allele of rs12913965 is associated with increased expression of the ISG20 gene compared to the C (reference) allele [7].

There are 71 genetic variants that are linked to rs12913965 in one linkage disequilibrium block spanning about 225 kb on chromosome 15 (using $R^2 > 0.7$ as a cutoff) (▶ **Fig. 3**, Supplemental **Table S1**). In principle, any of these variants might be responsible for the genetic association with shoulder dislocation with the remaining variants acting as neutral polymorphisms. None of these 71 SNPs affect the protein-coding capacity in the genes in this region. Besides the sentinel SNP rs12913965, none of the other SNPs show a strong association with shoulder dislocation; rs4932234 has the second strongest association in this linkage block with a p-value of 4.7×10^{-7}. In summary, the sentinel SNP rs12913965 has the best evidence for being responsible for the association with shoulder dislocation as it has by far the strongest p-value. This SNP is located within the TICRR gene and may affect the activity of this gene by mechanisms that are not yet understood, or it may associate with shoulder dislocation due to its effect on expression of ISG20, a gene located 1 Mb away.

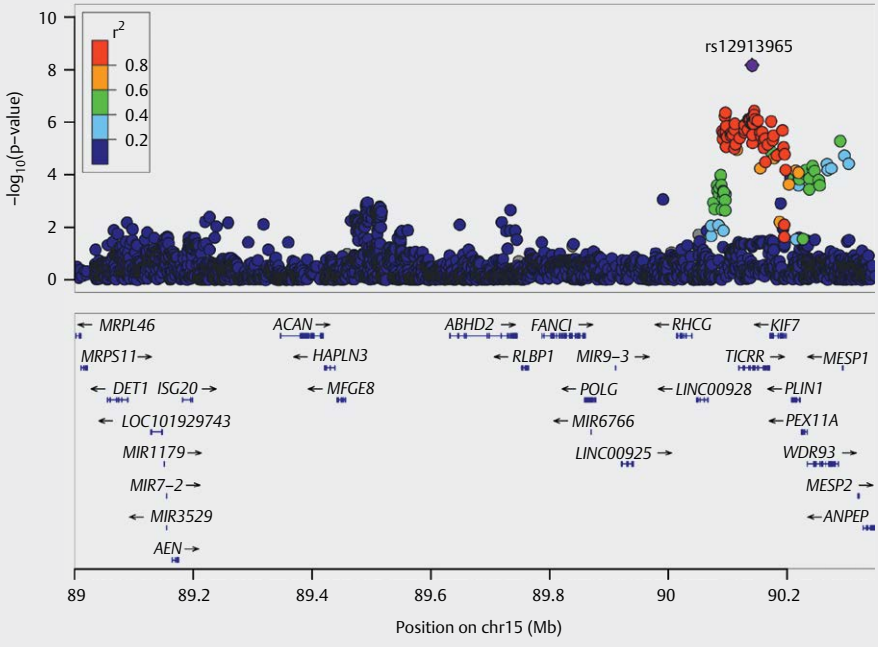
A previous study reported an association for rs180012 in the COL1A1 with shoulder dislocation and ACL rupture [20]. rs180012 was not represented amongst the 8,046,964 genotyped and imputed SNPs in our dataset, so we were not able to validate its association with shoulder dislocation.

Although the cases of shoulder dislocation for the LAT and EAS ancestry groups were too small in number to be informative, we performed a meta-analysis for the sake of inclusiveness. We first combined the data from the EUR and LAT logistic regressions in a 2-way meta-analysis. In the combined analysis, rs12913965 had a p-value of 5.5×10^{-8}, which is nearly genome-wide significant (Supplemental **Table S3**). However, nearly all of the statistical effect for the meta-analysis derives from the EUR ancestry group, as the p-value for rs12913965 in LAT was 0.9 (OR = 1.03; 95 % CI = 0.68–1.54). In the combined analysis with EUR, LAT and EAS, rs12913965 had a p-value of 3.8×10^{-7} (Supplemental **Table S3**). The p-value for rs12913965 in EAS was 0.64 (OR = 0.89; 95 % CI = 0.54–1.46).

▶ **Table 4** Genotype distributions for rs12913965 in the EUR ancestry group.

rs12913965	T/T	T/C	C/C
Cases	10	160	432
Controls	791	13 374	61 489
Overall	891	13 534	61 921
Risk for shoulder dislocation	1.12 %	1.18 %	0.70 %
Relative risk for shoulder dislocation (95 % CI)[a]	1.61 (0.86–3.02)[b]	1.69 (1.41–2.03)	1.000

[a] Risk for shoulder dislocation given that genotype compared to the risk given the C/C genotype; [b] Not significantly different than 1.000 (p = 0.10)



▶ **Fig. 3** Regional-association plot for rs12913965 with shoulder dislocation. Tested SNPs are arranged by genomic position on chromosome 15 (x-axis) in a 1.3 Mb window around the lead SNP rs12913965 (purple diamond). The y-axis indicates -$\log_{10}$ p-values for association with shoulder dislocation for each SNP in the EUR ancestry group. rs12913965 is located in an intron of the TICRR gene.

Discussion

Shoulder dislocations are a concern for athletes competing in contact sports such as football, rugby, wrestling and hockey [1, 11, 19, 35]. Very little is known about genetic risks for shoulder dislocation. A previous candidate gene study reported an association between rs180012 in col1A1 and cases with either ACL rupture or shoulder dislocation [20]. By obtaining access to large-scale genotype and phenotype data from the RPGEH, we were able to perform the first genome-wide screen for shoulder dislocation, and identified rs12913965 with a significant association in the EUR ancestry group. The data contained information from 83 264 EUR individuals including 662 cases of shoulder dislocation. Our results suggest that genetic differences affect an individual's intrinsic risk for shoulder dislocation.

Power calculations indicate that a cohort of this size would have about an 83 % chance of detecting a SNP with an association to MCL injury at genome-wide significance (assuming genotype relative risk of 2.3, minor allele frequency of 5 %). The chance of detection drops to 1 % for a SNP with a weaker effect on MCL injury (genotype relative risk of 1.5, minor allele frequency of 5 %).

As noted in our previous studies of genetic risk for injuries common in sporting activities using this cohort, our study included people regardless of whether or not they participated in a sport [21, 28, 29]. It is unknown whether the statistical association of rs12913965 with shoulder dislocation derives predominantly from the subset of the population that is active in one or more sports. It is also possible that this SNP acts by increasing a known risk factor, such as joint laxity [24].

Variation at rs12913965 may affect the risk for shoulder dislocation by affecting the activity of either TICRR or ISG20, which are 2 genes located nearby on chromosome 15. For the TICRR gene, rs12913965 is located within an intron of this gene but analyses have been unable to show that rs12913965 is linked to changes in its expression [7]. For the ISG20 gene, rs12913965 is located about 1 Mb away and variation in this SNP is associated with changes in its expression; specifically, the (T) allele is associated with higher expression of the ISG20 gene and increased risk for shoulder dislocation. The TICRR gene is involved in initiating DNA replication during the cell cycle and the ISG20 gene is involved in degrading viral RNA or DNA in response to interferon stimulation [9, 30]. Hence, a mechanistic link between these genes and shoulder dislocation is currently unclear.

For rs12913965, about 18 % of individuals were heterozygous for a risk allele (T/C genotype) and 81 % were homozygous for the protective allele (C/C genotype). Individuals carrying one risk allele (T) had a 69 % increased risk for shoulder dislocation compared to individuals with no risk alleles in our cohort.

Neither the LAT nor EAS ancestry groups showed an association for rs12913965 with shoulder dislocation. Power calculations show that, with a genotype relative risk of 1.6, our sentinel SNP would have a 50 % chance of validation at alpha = 0.05 for LAT (76 cases) and a 33 % chance of validation for EAS (45 cases). Hence, the non-significant results for rs12913965 with shoulder dislocation in LAT and EAS may reflect a small number of cases, rather than a lack of association in these ancestry groups. Larger cohort sizes are necessary to determine whether rs12913965 shows an association with shoulder dislocation in the LAT and EAS ancestry groups. Furthermore, we should acknowledge that when all the analyzable ancestry groups were included in meta-analyses, the detected signal was no longer formally genome-wide significant. Thus the EUR-specific signal needs to be viewed with caution until validated in other cohorts. Nevertheless, the p-value from the meta-analysis was still sufficiently close to genome-wide significance and empirical evidence suggests that p-values in this vicinity usually represent true signals [25].

As noted with our previous analyses using this cohort, one limitation of this type of study is that the phenotypes were defined from codes contained in the electronic health records, which may be inaccurate [21, 28, 29]. Another limitation is that the association between rs12913965 and shoulder dislocation has not yet been replicated in an independent cohort.

In the future, it will be interesting to perform the analysis on a population of athletes competing in sports that put them at risk for shoulder dislocation, such as football or basketball players. The results from these studies may reveal that rs12913965 could be used as a diagnostic marker to predict which athletes are at higher risk.

Conclusion

A genome-wide association screen revealed rs12913965 with a significant association for shoulder dislocation in the EUR ancestry group. Our data indicate that rs12913965 could explain part of the variation in risk for shoulder dislocation between individuals. In the future, sports medicine practitioners may be able to use this information to identify individuals with increased risk, and then take measures with stretching and strengthening exercises to prevent this injury before it occurs.

What are the findings?

- We performed a genome-wide association study for shoulder dislocation using data from Kaiser Permanente Northern California consisting of 662 cases and 82 602 controls of European descent.
- We discovered rs12913965 to be associated with shoulder dislocation at genome-wide significance in the EUR ancestry group ($p = 9.7 \times 10^{-9}$; OR = 1.6). Replication of this association in cohorts of athletes is warranted.

Acknowledgements

The authors thank the Kaiser Permanente Northern California RPGEH team for access to data and assistance in data management. Data used in this study were provided by the Kaiser Permanente Research Program on Genes, Environment, and Health (RPGEH): Genetic Epidemiology Research on Adult Health and Aging (GERA), funded by the National Institutes of Health [RC2 AG036607 (Schaefer and Risch)], the Robert Wood Johnson Foundation, the Wayne and Gladys Valley Foundation, The Ellison Medical Foundation, and the Kaiser Permanente Community Benefits Program. Access to RPGEH data used in this study may be obtained by application to the Kaiser Permanente Research Bank (KPRB) via ResearchBankAccess@kp.org. A subset of the GERA cohort consented for public use can be found at NIH/dbGaP: phs000674.v1.p1.

Conflict of interest

The authors have no conflict of interest to declare.

References

[1] Bohu Y, Klouche S, Lefevre N, Peyrin JC, Dusfour B, Hager JP, Ribaut A, Herman S. The epidemiology of 1345 shoulder dislocations and subluxations in French Rugby Union players: A five-season prospective study from 2008 to 2013. Br J Sports Med 2015; 49: 1535–1540

[2] Bottoni CR, Wilckens JH, DeBerardino TM, D'Alleyrand JC, Rooney RC, Harpstrite JK, Arciero RA. A prospective, randomized evaluation of arthroscopic stabilization versus nonoperative treatment in patients with acute, traumatic, first-time shoulder dislocations. Am J Sports Med 2002; 30: 576–580

[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] Burkart AC, Debski RE. Anatomy and function of the glenohumeral ligaments in anterior shoulder instability. Clin Orthop Relat Res 2002;http://www.ncbi.nlm.nih.gov/pubmed/12072743 32–39

[5] Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: Rising to the challenge of larger and richer datasets. Gigascience 2015; 4: 7

[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] Devlin B, Roeder K. Genomic control for association studies. Biometrics 1999; 55: 997–1004

[9] Espert L, Degols G, Gongora C, Blondel D, Williams BR, Silverman RH, Mechti N. ISG20, a new interferon-induced RNase specific for single-stranded RNA, defines an alternative antiviral pathway against RNA genomic viruses. J Biol Chem 2003; 278: 16151–16158

[10] Harriss DJ, Atkinson G. Ethical Standards in Sport and Exercise Science Research: 2016 Update. Int J Sports Med 2015; 36: 1121–1124

[11] Hazmy CH, Parwathi A. Sports-related shoulder dislocations: A state-hospital experience. Med J Malaysia 2005; 60 (SupplC): 22–25

[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] 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

[19] Kerr ZY, Collins CL, Pommering TL, Fields SK, Comstock RD. Dislocation/separation injuries among US high school athletes in 9 selected sports: 2005–2009. Clin J Sport Med 2011; 21: 101–108

[20] Khoschnau S, Melhus H, Jacobson A, Rahme H, Bengtsson H, Ribom E, Grundberg E, Mallmin H, Michaelsson K. Type I collagen alpha1 Sp1 polymorphism and the risk of cruciate ligament ruptures or shoulder dislocations. Am J Sports Med 2008; 36: 2432–2436

[21] Kim SK, Roos TR, Roos AK, Kleimeyer JP, Ahmed MA, Goodlin GT, Fredericson M, Ioannidis JPA, Avins AL, Dragoo JL. Genome-wide association screens for achilles tendon and ACL tears and tendinopathy. PLoS One 12(3): e0170422. doi:10.1371/journal.pone.0170422

[22] Kvale MN, Hesselson S, Hoffmann TJ, Cao Y, Chan D, Connell S, Croen LA, Dispensa BP, Eshragh J, Finn A, Gollub J, Iribarren C, Jorgenson E, Kushi LH, Lao R, Lu Y, Ludwig D, Mathauda GK, McGuire WB, Mei G, Miles S, Mittman M, Patil M, Quesenberry CP Jr., Ranatunga D, Rowell S, Sadler M, Sakoda LC, Shapero M, Shen L, Shenoy T, Smethurst D, Somkin CP, Van Den Eeden SK, Walter L, Wan E, Webster T, Whitmer RA, Wong S, Zau C, Zhan Y, Schaefer C, Kwok PY, Risch N. Genotyping informatics and quality control for 100,000 subjects in the Genetic Epidemiology Research on Adult Health and Aging (GERA) cohort. Genetics 2015; 200: 1051–1060

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

[24] Olds M, Ellis R, Donaldson K, Parmar P, Kersten P. Risk factors which predispose first-time traumatic anterior shoulder dislocations to recurrent instability in adults: A systematic review and meta-analysis. Br J Sports Med 2015; 49: 913–922

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

[26] 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

[27] Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, Sham PC. PLINK: A tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007; 81: 559–575

[28] Roos AK, Avins AL, Ahmed MA, Kleimeyer JP, Roos TR, Fredericson M, Ioannidis JPA, Dragoo JL, Kim SK. Two genetic loci associated with medial collateral ligament injury. Int J Sports Med.

[29] Roos TR, Roos AK, Avins AL, Ahmed MA, Kleimeyer JP, Frederidson M, Ioannidis JP, Dragoo JL, Kim SK. Genome-wide association study identifies a locus near cadherin8 associated with rotator cuff injury. submitted, DOI: http://dx.doi.org/10.1055/s-0043-104853

[30] Sansam CG, Goins D, Siefert JC, Clowdus EA, Sansam CL. Cyclin-dependent kinase regulates the length of S phase through TICRR/TRESLIN phosphorylation. Genes Dev 2015; 29: 555–566

[31] Sansam CL, Cruz NM, Danielian PS, Amsterdam A, Lau ML, Hopkins N, Lees JA. A vertebrate gene, ticrr, is an essential checkpoint and replication regulator. Genes Dev 2010; 24: 183–194

[32] 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: 209–213

[33] te Slaa RL, Wijffels MP, Brand R, Marti RK. The prognosis following acute primary glenohumeral dislocation. J Bone Joint Surg Br 2004; 86: 58–64

[34] Ward LD, Kellis M. HaploReg v4: Systematic mining of putative causal variants, cell types, regulators and target genes for human complex traits and disease. Nucleic Acids Res 2016; 44: D877–D881

[35] Zacchilli MA, Owens BD. Epidemiology of shoulder dislocations presenting to emergency departments in the United States. J Bone Joint Surg Am 2010; 92: 542–549