Published in final edited form as:

J Bone Joint Surg Am. 2021 July 21; 103(14): 1259–1267. doi:10.2106/JBJS.20.01474.

Identification of a Novel Genetic Marker for Risk of Degenerative Rotator Cuff Disease Surgery in the UK Biobank

Elizabeth L. Yanik, PhD ScM[1,2], **Jay D. Keener, MD**[1], **Shiow J. Lin**[3], **Graham A. Colditz, DrPH MD**[2], **Rick W. Wright, MD**[1], **Bradley A. Evanoff, MD MPH**[4], **Nitin B. Jain, MD MSPH**[5], **Nancy L. Saccone, PhD**[3]

[1]Department of Orthopaedic Surgery, Washington University School of Medicine, St. Louis, MO

[2]Department of Surgery, Washington University School of Medicine, St. Louis, MO

[3]Department of Genetics, Washington University School of Medicine, St. Louis, MO

[4]Division of General Medical Sciences, Washington University School of Medicine, St. Louis, MO

[5]Department of Physical Medicine and Rehabilitation, University of Texas Southwestern, Dallas, TX

Abstract

BACKGROUND: While evidence indicates that familial predisposition influences degenerative rotator cuff disease (RCD) risk, knowledge of specific genetic markers is limited. We conducted a genome-wide association study of RCD surgery using UK Biobank, a prospective cohort of 500,000 people aged 40–69 at enrollment with genotype data.

METHODS: Degenerative RCD surgery cases were identified using linked hospital records. Cases were defined as presence of ICD-10 code M75.1 diagnosed by a trauma/orthopedic specialist with accompanying surgery consistent with RCD treatment. Cases were excluded if traumatic injury diagnoses were made during the same hospital visit. For each case up to five controls were chosen from UK Biobank matched by age, sex, and follow-up time. Analyses were limited to European-ancestry individuals who were not third degree or closer. We used logistic regression to test for genetic association of 674,405 typed and >10 million imputed markers adjusting for age, sex, population principal components, and follow-up.

RESULTS: We identified 2,917 RCD surgery cases and 14,158 matched controls. We observed one genome-wide significant signal (p-value$<5\times10^{-8}$) for a novel locus tagged by rs2237352 in the CREB5 gene on chromosome 7 (OR=1.17, 95%CI=1.11–1.24). Single nucleotide polymorphism (SNP) rs2237352 was imputed with a high degree of confidence (info score=0.9847) and is common with a minor allele frequency of 47%. After expanding the control

Corresponding Author: Elizabeth L. Yanik, Assistant Professor, Department of Orthopaedic Surgery, Division of Public Health Sciences, Department of Surgery, Washington University School of Medicine, Phone: 314-747-3196, yanike@wustl.edu.
Institution Addresses: Washington University School of Medicine, 660 S. Euclid Ave., St. Louis, MO 63110, Vanderbilt University School of Medicine, 1161 21st. Ave., Nashville, TN 37232, University of Texas Southwestern, 5323 Harry Hines Blvd., Dallas, TX 75390

Author Contributions: ELY developed the initial study design and drafted the manuscript. ELY, SJL, and NLS conducted the statistical analysis. GAC, RWW, NLS, BAE, NBJ, and JDK contributed to the study design, interpretation of the results, and revisions to the manuscript. All authors approved the final manuscript.

sample to include additional unmatched non-cases, rs2237352 and another SNP on the CREB5 gene, rs12700903, were genome-wide significant. We did not detect genome-wide significant signals at loci associated with RCD in previous studies.

CONCLUSIONS: We identified a novel association between a variant in the CREB5 gene and RCD surgery. Validation of this finding in studies with imaging data to confirm diagnoses will be important.

CLINICAL RELEVANCE: Identification of genetic RCD susceptibility markers can guide understanding of biological processes in cuff degeneration and help inform disease risk in the clinical setting.

Introduction

Rotator cuff disease (RCD) is the most common cause of shoulder disability,[1–4] but studies identifying risk factors for symptomatic disease have been sparse. Prevention strategies aimed at high risk groups could dramatically impact healthcare given the economic burden of treating painful RCD.[5,6]

Accumulating evidence indicates that familial predisposition influences degenerative RCD risk.[7–10] However, knowledge of specific RCD genetic markers is limited. Studies have evaluated candidate genes and discovered associations with genes involved in repair and degenerative processes, including genes expressed in response to tissue damage.[11–13] The first genome wide association study (GWAS) for RCD detected two associated single nucleotide polymorphisms (SNPs) involved in apoptosis,[14] but this study of <350 RCD patients had limited statistical power. A second GWAS was unable to replicate associations from the first, and identified a new SNP associated with RCD.[15] While this study was much larger (8,357 rotator cuff injury cases), the definition for RCD was non-specific, using ICD codes that might capture shoulder pain from other diseases.[15]

Uncertainty still exists about which genetic markers have true associations with degenerative RCD. The UK Biobank population of half a million people with genotype data provided a unique opportunity for a large GWAS with carefully defined RCD cases and controls to identify additional genetic markers and further evaluate the replicability of previous findings.

Methods

Our study population was derived from UK Biobank, a population-based prospective cohort of approximately 500,000 United Kingdom residents.[16,17] Participants aged 40–69 years were recruited nationwide from 2006–2010 through invitations mailed to people registered with the National Health Service (NHS).[18] At enrollment participants gave informed consent and whole blood samples were collected. NHS hospital records were linked to UK Biobank providing information on inpatient diagnoses and procedures during 2006–2017. Diagnoses were coded using the International Classification of Diseases, 10th revision (ICD-10). Procedures were coded using the Office of Population Censuses and Surveys Classification, 4th revision (OPCS-4). DNA was extracted from whole blood samples and genotyped using

the UK Biobank Axiom Array, which includes 812,428 SNPs and insertion-deletion markers.[17,19] An additional 73 million markers were imputed using a reference haplotype panel to predict genetic markers not directly assayed. We obtained de-identified data from UK Biobank (project number 27034) on 488,292 UK Biobank participants with available genotype results. Of these, 968 participants were excluded because of poor quality results indicated by either extreme heterozygosity or missingness.[17]

Within the remaining 487,324 people we selected cases and controls. As >94% of participants self-reported European/white ancestry, the case-control population was limited to this group to reduce population stratification. Numerous quality control (QC) measures were undertaken to confirm data quality of the selected case-control sample.[20,21] People were excluded if there was discordance between their reported and genetic sex, or a typed autosomal call rate <0.95. Population principal components (PCs) were used to infer shared ancestry.[22] Individuals who reported European/white ancestry, but whose PCs did not cluster with the British ancestry cohort were excluded.[17] We included individuals whose first three PCs were near the mean of this British cluster, namely within one third the full range of that PC across populations. Related pairs up to a third degree were provided by UK Biobank. To facilitate GWAS methods that assume unrelated individuals, one person from each related pair was excluded. When a pair contained a case (as defined below) and a non-case, the non-case was preferentially excluded. Otherwise, one person was excluded at random. The remaining subcohort formed the population from which cases and controls were selected.

Degenerative RCD cases were identified based on a primary or secondary ICD-10 diagnosis code of M75.1. To reduce risk of misdiagnosis, RCD cases were included only if a diagnosis was made by a trauma/orthopedics specialist and an accompanying surgical procedure consistent with RCD treatment was present. As genetic factors are considered important primarily for degenerative RCD, cases were excluded if they occurred concurrently with traumatic injury diagnoses (ex. shoulder dislocation).

Five controls were randomly selected for each case through incidence-density sampling, in which controls are randomly chosen from individuals in follow-up without a prior degenerative RCD diagnosis at the time of case diagnosis. Incidence-density sampling ensures a representative sample population with comparable follow-up for cases and controls.[23,24] Follow-up was defined as time since enrollment in UK Biobank. Controls were matched to cases by age and reported sex.

The NHS Research Ethics Committee approved the UK Biobank. The Washington University Institutional Review Board determined this study to be exempt from oversight.

We included genetic variants that did not diverge from Hardy-Weinberg equilibrium with a p-value$<1\times10^{-6}$ and that had a minor allele frequency (MAF) >0.004285. The MAF cut-off was based on the formula 25/(2*number of cases) so ≥25 minor alleles would be expected in cases under the null hypothesis. Only imputed variants with an INFO score ≥0.3 were included to remove variants with low confidence imputed values.

Association tests used logistic regression to model genotype dosage effects on RCD, with covariates for age, sex, follow-up time, and the first 10 population PCs.[22] QQ-plots were

graphed and a genomic inflation factor calculated to check for bias. We required a genome-wide significance threshold of 5×10^{-8}.[20,25] For regions harboring GWAS significant signals, we performed an adjusted analysis using the lead SNP as a covariate to detect additional independent signals. We also specifically examined genetic markers identified as significantly associated with RCD in prior literature.[11–15,26–30] For these markers a Bonferroni-adjusted p-value<0.05 and an odds ratio (OR) indicating an association in the same direction as the original publication were considered evidence of replication. PLINK (v1.9 and v2.0) was used for data cleaning and analysis.[31,32]

For genome-wide significant variants identified in initial analyses, several sensitivity analyses were undertaken. First, associations were estimated using the same cases and a larger set of controls defined as all non-cases in UK Biobank who met the quality control standards outlined above. Second, associations were estimated using conditional logistic regression to incorporate individuals selected multiple times as controls through incidence-density sampling and more precisely adjust for matching criteria. Third, associations were estimated in the subgroup of cases ≤60 years old at RCD surgery (and corresponding controls), as we hypothesized that genetic predisposition could lead to earlier-onset disease.

Results

Of the 487,324 people with usable genotype information, 28,072 people reported non-white ancestry (Figure 1). We excluded 364 people with a typed autosomal call rate <0.95, 357 people with discordance between reported and genetic sex, and 78,859 people to eliminate relatedness up to the 3rd degree. Thirty-one people were excluded based on population PCs.

In the remaining 379,641 people, we identified 2,917 degenerative RCD surgery cases. For cases, diagnoses occurred after a median 5 years of follow-up (interquartile range=3–6 years, range=0–10 years). The median diagnosis age was 65 years old (interquartile range=59–69, range=41–78), and 48% of cases were women (Table 1). Cases were follow-up, age, and sex-matched with a 1:5 ratio to 14,158 unique controls. For conditional logistic regression to represent incidence-density sampling, some individuals could be selected as controls multiple times, giving 14,547 controls.

Initially, over 77 million typed and imputed variants were available for analyses. Of these, >66 million were removed because of MAF<0.004285 and 50,998 were removed because of a Hardy-Weinberg exact test p-value$<1\times10^{-6}$. There remained 674,405 typed and 10,140,917 imputed variants included in analyses.

The Q-Q plot and genomic inflation factor of 1.02 provided no evidence for bias after accounting for matching factors and the first 10 PCs (Figure 2).

We observed one novel genome-wide significant signal (p-value=4.04×10^{-8}) at SNP rs2237352 in the CREB5 gene on chromosome 7 (OR=1.17, 95%CI=1.11–1.24, Table 2, Figure 3). SNP rs2237352 was imputed with a high degree of confidence (INFO score=0.9847) and is a common variant (MAF=46.8%). The second strongest signal was for imputed SNP rs12700903 in the CREB5 gene (OR=1.17, 95%CI=1.11–1.24, p-value=5.63×10^{-8}), which is in strong linkage disequilibrium (LD) with rs2237352 (r^2 =

0.98). Thus both SNPs represent the same statistical signal. The most significant directly assayed SNP in the CREB5 gene was rs66539057, but the association was not significant at a genome-wide level (OR=1.16, 95%CI=1.09–1.23, p-value=1.29×10^{-6}). Figure 4 shows a detailed view of the associated region, with the lead SNP rs2237352 having strong to moderate LD with additional SNPs in the region.[33] The strongest signal for a SNP having modest LD with rs2237352 was for rs4722837 (OR=0.86, p-value=1.26E-7, r^2=0.38 with rs2237352).

Results were similar in conditional logistic regression (rs2237352 OR=1.17, 95%CI=1.10–1.24; rs12700903 OR=1.17, 95%CI=1.10–1.24, Table 2).

After analyses adjusting for rs2237352 genotype, no additional SNPs in the region showed strong association (lowest p-value within 500Kb was 1.5×10^{-4}).

We did not detect genome-wide significant signals at 28 loci associated with RCD in other studies. Two SNPs, rs820218 in the SAP30BP gene and rs2277698 at the TIMP2 gene, were associated with RCD surgery with a nominal p-value<0.05 (rs820218 OR=0.93, 95%CI=0.88–0.99, p-value=0.025; rs2277698 OR=0.87, 95%CI=0.80–0.95, p-value=0.002, Table 3). However, for rs2277698 the association was in the opposite direction as in the prior study. After Bonferroni-adjustment for 28 replication attempts, neither SNP remained significant.

After expanding controls to include the larger, unmatched cohort of non-cases (N=375,560), rs2237352 remained genome-wide significant (p-value=2.29×10^{-8}, Table 2). Additionally, the rs12700903 association with degenerative RCD surgery became genome-wide significant (p-value=3.69×10^{-8}).

There were 735 cases ≤60 years old at diagnosis and 3602 corresponding controls. Within this subgroup, associations with rs2237352 and rs12700903 were not genome wide-significant due to reduced statistical power. Odds ratios were slightly larger (rs2237352 OR=1.22, 95%CI=1.09–1.37, p-value=0.0006; rs12700903 OR=1.23, 95%CI=1.09–1.38, p-value=0.0005).

Discussion

We conducted a large, population-based GWAS of degenerative RCD surgery with well-defined cases and identified a novel association with a common SNP in the CREB5 gene. SNP associations from prior RCD studies did not replicate in our population after multiple test correction. These findings highlight the need for further large GWAS of degenerative RCD focused on carefully defining case and control status and identifying replicable results.

We identified a novel association between SNP rs2237352 and degenerative RCD surgery. SNP rs12700903 was also associated with degenerative RCD surgery after expansion of our control group, and represents the same signal as SNP rs2237352. Both SNPs are located on the CREB5 gene, which encodes a protein that is part of the cAMP response element-binding protein family.[34] CREB5 is a transcription factor involved in cell growth, proliferation, and differentiation.[35,36] CREB5 expression has been associated with plasma

interleukin-6 levels and may influence inflammatory response genes.[37] As CREB family proteins influence expression of other genes,[38,39] there may be numerous genetic mutations that could influence the same biologic pathways. If further research confirms this association, one would expect genetic risk for RCD to be highly polygenic as is common for most complex traits. Differential CREB5 expression has also been specifically documented in fibroblasts,[40] lending further evidence that mutations in this gene could be of importance for tendon injury and repair.

After adjusting for rs2237352 in models, no additional signals were detected, consistent with this region harboring one primary locus associated with degenerative RCD. CREB5 SNPs in weaker LD with the top signal did not provide GWAS-significant evidence for another distinct signal in the region. However, this locus could represent an accumulation of weak effects from linked variants that influence degenerative RCD.[41] Notably, as rs2237352 is an intron variant, it may be indicative of an unknown genetic determinant with which it co-segregates.

Most prior RCD genetic epidemiology studies have been candidate gene studies, while two GWAS RCD studies in independent populations have been conducted.[10,13,15] Candidate gene studies focus on specific genes with known function potentially related to rotator cuff degeneration, whereas GWAS studies take an agnostic approach to testing association with large portions of the genome. Of the 28 SNPs from prior studies that we could evaluate in UK Biobank, only 1 demonstrated an association in UK Biobank in the same direction as the prior study with a nominal, uncorrected $P<0.05$ (rs820218), while none reached genome-wide significance. SNP rs820218 is located on the SAP30BP gene, which encodes a transcriptional regulator protein involved in cell death and apoptosis.[14,42] Numerous studies have shown increased tendon cell apoptosis related to rotator cuff tearing.[43–45] However, as only 1 SNP out of 28 demonstrated a consistent association, this could indicate a chance replication.

Other findings were not replicated in our study. These discrepancies could be due to initial findings being due to chance, or could result from differences in case definitions. As our cases were specifically drawn from surgical cases in a hospital, they likely represent more severe disease than cases in studies capturing all confirmed rotator cuff tears in an orthopaedic clinical setting. For SNPs reported from candidate gene studies, we were unable to replicate associations at the $P<0.05$ level. Candidate gene studies have had limited success in identifying large portions of the genetic contributions to other complex diseases, and can be more susceptible to publication bias.[46–48]

One other genetic RCD study has recently been conducted in UK Biobank, which used a less specific case definition that appeared to include primary care diagnoses without treatment information.[49] The associated SNPs in our study were not identified as having genome-wide significant signals in that study. Instead, three other associated SNPs were identified. These SNPs in the GLCCI1, THSD7A, and ZNF804A genes showed similar associations in our results that were not genome-wide significant (ex. for variant in GLCC1 gene: odds ratio of 0.88 vs. 0.89 in our study with P-value=2.84×10^{-4}). The contrasting results likely reflect differing case definitions. The other study's less stringent definition

could capture cases that represent alternate sources of shoulder pain given preliminary diagnoses of RCD, while we expect our cases to disproportionately capture severe disease.

There are important limitations to the current study. First, we could not confirm rotator cuff tears with MRI or ultrasound as shoulder imaging was not available within UK Biobank. Relatedly, our control group could include people with asymptomatic tears or symptomatic tears being treated conservatively or in an outpatient setting. The prevalence of shoulder conditions in the UK primary care setting is reported to be 16–26% with most relating to rotator cuff tendon problems.[50,51] We also could not differentiate between genetic effects that influence tear initiation, progression in tear severity, and pain tolerance. We did not have information on treatment outcomes and so could not evaluate how genetics influenced surgical outcomes. As is usual for GWAS results, the associated SNPs might not be directly causal, but instead merely correlated with functionally significant variants. Finally, as our analyses were limited to the white/European-ancestry population, further studies in non-European ancestry populations will be important in the future to examine the generalizability of our results.

Our study also had numerous strengths. This is one of the largest genetic RCD studies that has been conducted to date.[15] Our incorporation of diagnostic, procedure, and provider specialty information likely made our case definition more specific than other large RCD GWAS,[15,49] and we did not appear to be substantially under-capturing RCD surgery cases based on rates in the literature.[52] A large number of genetic markers were available for examination, including typed markers and markers imputed with a high degree of confidence. This strength of UK Biobank will improve further in the future as plans are in place for whole genome sequencing of the population.[53]

A more comprehensive understanding of genetic susceptibility for degenerative RCD could aid treatment and prevention in several ways. First, someone with a genetic predisposition for RCD could derive greater benefits from changing modifiable risk factors such as smoking or occupational burdens. Second, a predisposition for cuff degeneration may also indicate impaired ability of the cuff to heal following surgical repair, which could influence cuff repair indications. Third, genetic susceptibility markers may point to key biological pathways in cuff degeneration that could direct future basic science research leading to novel therapeutics.

We identified a novel SNP in the CREB5 gene associated with degenerative RCD surgery in a general population sample of the United Kingdom. Further replication of this finding will be important in the future. Future examination of the genetic determinants of other chronic tendon disorders, including investigation of commonalities across such disorders, would be useful. The extensive information available in UK Biobank could allow future evaluation of risk models incorporating genetics, non-genetic characteristics, and gene-environment interactions.[54] Identification of potentially important genetic markers in our study and others can allow a more focused study of these markers in smaller cohorts with more detailed clinical information, including investigations of how genetic factors may influence RCD progression and outcomes after surgical treatment.

ACKNOWLEDGEMENTS:

ELY received funding from the National Institutes of Health K01AR073318 and the Orthopaedic Research and Education Foundation/American Society of Shoulder and Elbow Surgeons Rockwood Clinical Research Grant in Shoulder Care. BAE received funding from the US National Institute for Occupational Safety and Health R01 OH011076. This research has been conducted using the UK Biobank Resource (http://www.ukbiobank.ac.uk/). Returned results, including GWAS association results for all typed and imputed variants tests, can be found under application 27034.

Funding

This investigation was supported by the National Institute of Arthritis and Musculoskeletal and Skin Diseases, K01AR073318.

References

1. Yamamoto A, Takagishi K, Osawa T, et al. Prevalence and risk factors of a rotator cuff tear in the general population. J Shoulder Elbow Surg 2010; 19(1): 116–20. [PubMed: 19540777]
2. Liem D, Buschmann VE, Schmidt C, et al. The prevalence of rotator cuff tears: is the contralateral shoulder at risk? Am J Sports Med 2014; 42(4): 826–30. [PubMed: 24500916]
3. Sher JS, Uribe JW, Posada A, Murphy BJ, Zlatkin MB. Abnormal findings on magnetic resonance images of asymptomatic shoulders. J Bone Joint Surg Am 1995; 77A(1): 10–5.
4. Tempelhof S, Rupp S, Seil R. Age-related prevalence of rotator cuff tears in asymptomatic shoulders. J Elbow Shoulder Surg 1999; 8: 296–9.
5. Mather RC 3rd, Koenig L, Acevedo D, et al. The societal and economic value of rotator cuff repair. J Bone Joint Surg Am 2013; 95(22): 1993–2000. [PubMed: 24257656]
6. Colvin AC, Egorova N, Harrison AK, Moskowitz A, Flatow EL. National trends in rotator cuff repair. J Bone Joint Surg Am 2012; 94(3): 227–33. [PubMed: 22298054]
7. Tashjian RZ. Epidemiology, natural history, and indications for treatment of rotator cuff tears. Clin Sports Med 2012; 31(4): 589–604. [PubMed: 23040548]
8. Hsu J, Keener JD. Natural History of Rotator Cuff Disease and Implications on Management. Oper Tech Orthop 2015; 25(1): 2–9. [PubMed: 26726288]
9. Tashjian RZ, Farnham JM, Granger EK, Teerlink CC, Cannon-Albright LA. Evidence for an Environmental and Inherited Predisposition Contributing to the Risk for Global Tendinopathies or Compression Neuropathies in Patients With Rotator Cuff Tears. Orthop J Sports Med 2016; 4(4): 2325967116642173. [PubMed: 27115018]
10. Dabija DI, Gao C, Edwards TL, Kuhn JE, Jain NB. Genetic and familial predisposition to rotator cuff disease: a systematic review. J Shoulder Elbow Surg 2017; 26(6): 1103–12. [PubMed: 28162885]
11. Motta Gda R, Amaral MV, Rezende E, et al. Evidence of genetic variations associated with rotator cuff disease. J Shoulder Elbow Surg 2014; 23(2): 227–35. [PubMed: 24129055]
12. Bonato LL, Quinelato V, Pinheiro Ada R, et al. ESRRB polymorphisms are associated with comorbidity of temporomandibular disorders and rotator cuff disease. Int J Oral Maxillofac Surg 2016; 45(3): 323–31. [PubMed: 26584852]
13. Kluger R, Burgstaller J, Vogl C, Brem G, Skultety M, Mueller S. A candidate gene approach identifies six SNPs in tenascin-C (TNC) associated with degenerative rotator cuff tears. J Orthop Res 2016.
14. Tashjian RZ, Granger EK, Farnham JM, Cannon-Albright LA, Teerlink CC. Genome-wide association study for rotator cuff tears identifies two significant single-nucleotide polymorphisms. J Shoulder Elbow Surg 2016; 25(2): 174–9. [PubMed: 26350878]
15. Roos TR, Roos AK, Avins AL, et al. Genome-wide association study identifies a locus associated with rotator cuff injury. PLoS One 2017; 12(12): e0189317. [PubMed: 29228018]
16. Collins R What makes UK Biobank special? The Lancet 2012; 379(9822): 1173–4.
17. Bycroft C, Freeman C, Petkova D, et al. The UK Biobank resource with deep phenotyping and genomic data. Nature 2018; 562(7726): 203–9. [PubMed: 30305743]

18. Manolio TA, Weis BK, Cowie CC, et al. New models for large prospective studies: is there a better way? Am J Epidemiol 2012; 175(9): 859–66. [PubMed: 22411865]
19. Elliott P, Peakman TC, Biobank UK. The UK Biobank sample handling and storage protocol for the collection, processing and archiving of human blood and urine. Int J Epidemiol 2008; 37(2): 234–44. [PubMed: 18381398]
20. de Bakker PI, Ferreira MA, Jia X, Neale BM, Raychaudhuri S, Voight BF. Practical aspects of imputation-driven meta-analysis of genome-wide association studies. Hum Mol Genet 2008; 17(R2): R122–8. [PubMed: 18852200]
21. Marees AT, de Kluiver H, Stringer S, et al. A tutorial on conducting genome-wide association studies: Quality control and statistical analysis. Int J Methods Psychiatr Res 2018; 27(2): e1608. [PubMed: 29484742]
22. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 2006; 38(8): 904–9. [PubMed: 16862161]
23. Lubin JH, Gail MH. Biased selection of controls for case-control analyses of cohort studies. Biometrics 1984; 40(1): 63–75. [PubMed: 6375751]
24. Richardson DB. An incidence density sampling program for nested case-control analyses. Occup Environ Med 2004; 61(12): e59. [PubMed: 15550597]
25. Barsh GS, Copenhaver GP, Gibson G, Williams SM. Guidelines for genome-wide association studies. PLoS Genet 2012; 8(7): e1002812. [PubMed: 22792080]
26. Peach CA, Zhang Y, Dunford JE, Brown MA, Carr AJ. Cuff tear arthropathy: evidence of functional variation in pyrophosphate metabolism genes. Clin Orthop Relat Res 2007; 462: 67–72. [PubMed: 17563703]
27. Figueiredo EA, Loyola LC, Belangero PS, et al. Rotator Cuff Tear Susceptibility Is Associated With Variants in Genes Involved in Tendon Extracellular Matrix Homeostasis. J Orthop Res 2020; 38(1): 192–201. [PubMed: 31444797]
28. Kluger R, Huber KR, Seely PG, Berger CE, Frommlet F. Novel Tenascin-C Haplotype Modifies the Risk for a Failure to Heal After Rotator Cuff Repair. Am J Sports Med 2017; 45(13): 2955–64. [PubMed: 28952802]
29. Teerlink CC, Cannon-Albright LA, Tashjian RZ. Significant association of full-thickness rotator cuff tears and estrogen-related receptor-beta (ESRRB). J Shoulder Elbow Surg 2015; 24(2): e31–5. [PubMed: 25219474]
30. Tashjian RZ, Granger EK, Zhang Y, Teerlink CC, Cannon-Albright LA. Identification of a genetic variant associated with rotator cuff repair healing. J Shoulder Elbow Surg 2016; 25(6): 865–72. [PubMed: 27066960]
31. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007; 81(3): 559–75. [PubMed: 17701901]
32. Chang CC, Chow CC, Tellier LCAM, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience 2015; 4: s13742–015-0047–8.
33. Pruim JR, Welch RP, Sanna A, et al. LocusZoom: Regional Visualization of Genome-Wide Association Scan Results. Bioinformatics 2010; 26: 2336–7. [PubMed: 20634204]
34. National Center for Biotechnology Information. CREB5 cAMP responsive element binding protein 5 [Homo sapiens (human)]. Gene. https://www.ncbi.nlm.nih.gov/gene?cmd=Retrieve&dopt=full_report&list_uids=9586: National Center for Biotechnology Information,; 2020.
35. Yu M, Du G, Xu Q, et al. Integrated analysis of DNA methylome and transcriptome identified CREB5 as a novel risk gene contributing to recurrent pregnancy loss. EBioMedicine 2018; 35: 334–44. [PubMed: 30100398]
36. He S, Deng Y, Liao Y, Li X, Liu J, Yao S. CREB5 promotes tumor cell invasion and correlates with poor prognosis in epithelial ovarian cancer. Oncol Lett 2017; 14(6): 8156–61. [PubMed: 29250192]
37. Nevalainen T, Kananen L, Marttila S, et al. Transciptomic and epigenetic analyses reveal a gender difference in aging-associated inflammation: the Vitality 90+ study. Age 2015; 37(76).

38. Bourtchuladze R, Frenguelli B, Blendy J, Cioffi D, Schutz G, Silva AJ. Deficient Long-Term Memory in Mice with a Targeted Mutation of the cAMP-Responsive Element-Binding Protein. Cell 1994; 79(1): 59–68. [PubMed: 7923378]
39. Montminy MR, Bilezikjian LM. Binding of a nuclear protein to the cyclic-AMP response element of the somatostatin gene. Nature 1987; 328(6126): 175–8. [PubMed: 2885756]
40. The Broad Institute of MIT and Harvard. GTeX Portal. CREB5. https://www.gtexportal.org/home/: The Broad Institute of MIT and Harvard; 2019.
41. Boyle EA, Li YI, Pritchard JK. An Expanded View of Complex Traits: From Polygenic to Omnigenic. Cell 2017; 169(7): 1177–86. [PubMed: 28622505]
42. Li JF, Liu LD, Ma SH, et al. HTRP--an immediate-early gene product induced by HSV1 infection in human embryo fibroblasts, is involved in cellular co-repressors. J Biochem 2004; 136(2): 169–76. [PubMed: 15496587]
43. Benson RT, McDonnell SM, Knowles HJ, Rees JL, Carr AJ, Hulley PA. Tendinopathy and tears of the rotator cuff are associated with hypoxia and apoptosis. J Bone Joint Surg Br 2010; 92: 488–53.
44. Lee HJ, Kim YS, Ok JH, Song HJ. Apoptosis occurs throughout the diseased rotator cuff. Am J Sports Med 2013; 41(10): 2249–55. [PubMed: 23788682]
45. Osti L, Buda M, Del Buono A, Osti R, Massari L, Maffulli N. Apoptosis and rotator cuff tears: scientific evidence from basic science to clinical findings. Br Med Bull 2017; 122(1): 123–33. [PubMed: 28369181]
46. Ioannidis JP, Ntzani EE, Trikalinos TA, Contopoulos-Ioannidis DG. Replication validity of genetic association studies. Nat Genet 2001; 29(3): 306–9. [PubMed: 11600885]
47. Tabor HK, Risch NJ, Myers RM. Candidate-gene approaches for studying complex genetic traits: practical considerations. Nature Reviews Genetics 2002; 3(5): 391–7.
48. Border R, Johnson EC, Evans LM, et al. No Support for Historical Candidate Gene or Candidate Gene-by-Interaction Hypotheses for Major Depression Across Multiple Large Samples. Am J Psychiatry 2019; 176: 376–87. [PubMed: 30845820]
49. Tashjian RZ, Kim SK, Roche MD, Jones KB, Teerlink CC. Genetic Variants Associated With Rotator Cuff Tearing Utilizing Multiple Population-Based Genetic Resources. J Shoulder Elbow Surg 2020.
50. Urwin M, Symmons D, Allison T, et al. Estimating the burden of musculoskeletal disorders in the community: the comparative prevalence of symptoms at different anatomical sites, and the relation to social deprivation. Ann Rheum Dis 1998; 57: 649–55. [PubMed: 9924205]
51. Ostor AJ, Richards CA, Prevost AT, Speed CA, Hazleman BL. Diagnosis and relation to general health of shoulder disorders presenting to primary care. Rheumatology (Oxford) 2005; 44(6): 800–5. [PubMed: 15769790]
52. Judge A, Murphy RJ, Maxwell R, Arden NK, Carr AJ. Temporal trends and geographical variation in the use of subacromial decompression and rotator cuff repair of the shoulder in England. Bone Joint J 2014; 96-B: 70–4.
53. UK Biobank. UK Biobank leads the way in genetics research. 2019 (accessed 13 May 2020.
54. Yanik EL, Colditz GA, Wright RW, et al. Risk factors for surgery due to rotator cuff disease in a population-based cohort. Bone Joint J 2020; 102-B(3): 352–9. [PubMed: 32114822]

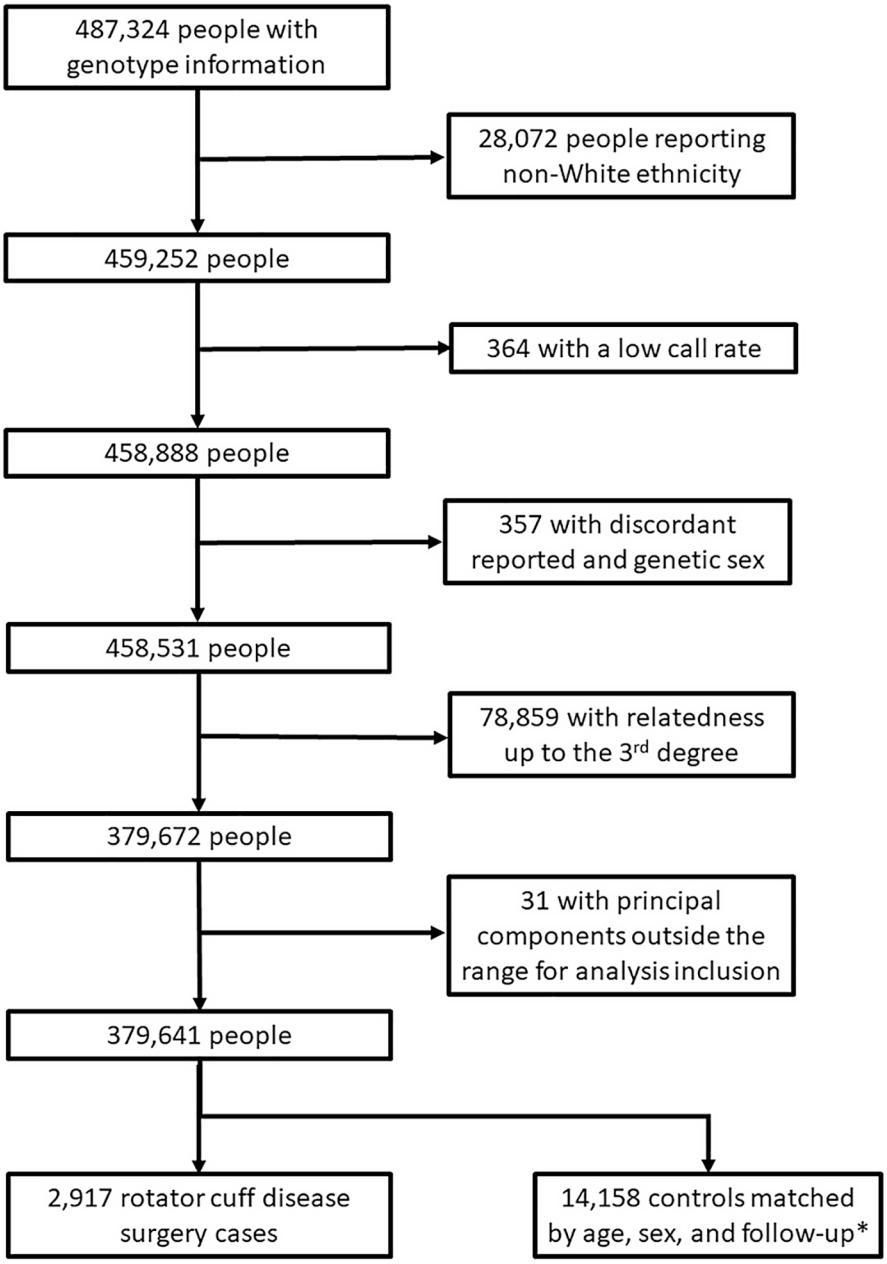

Figure 1. Flow chart of exclusions made prior to selection of rotator cuff surgery cases and controls in the UK Biobank

*Analyses using matched controls, such as conditional logistic regression, allowed individuals to serve as controls for multiple cases or serve as a case later in follow-up. As a result, 14,547 controls were identified for these analyses.

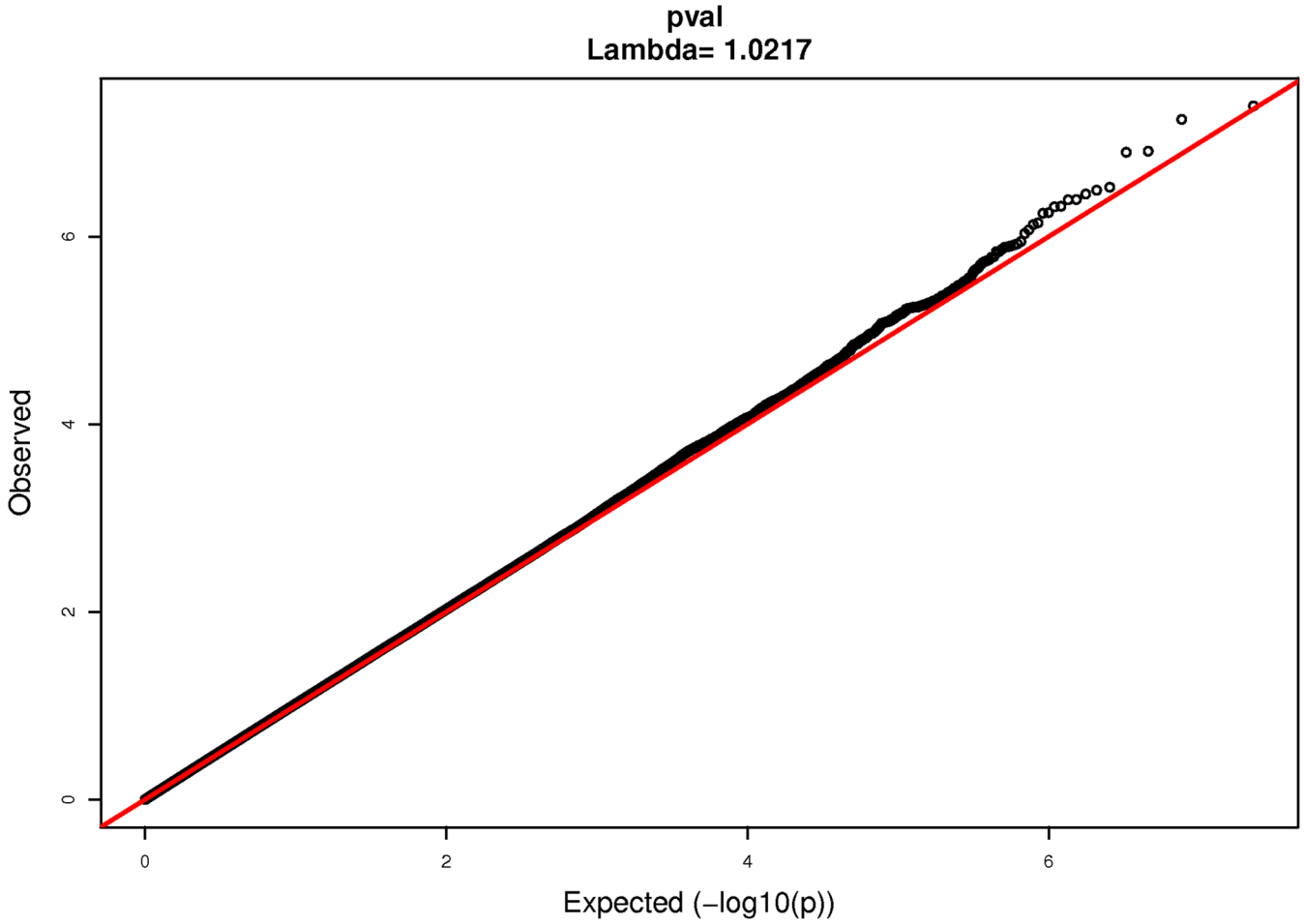

Figure 2. Quantile-Quantile plot comparing observed P-values to the expected distribution of P-values for each association of a genetic variant with rotator cuff disease surgery in the UK Biobank

Substantial, systematic divergence of the distribution of data points from the red diagonal line would indicate bias. Lambda represents the genomic inflation factor calculated by dividing the median observed test statistics by the median expected test statistic. A genomic inflation factor of 1 indicates no bias.

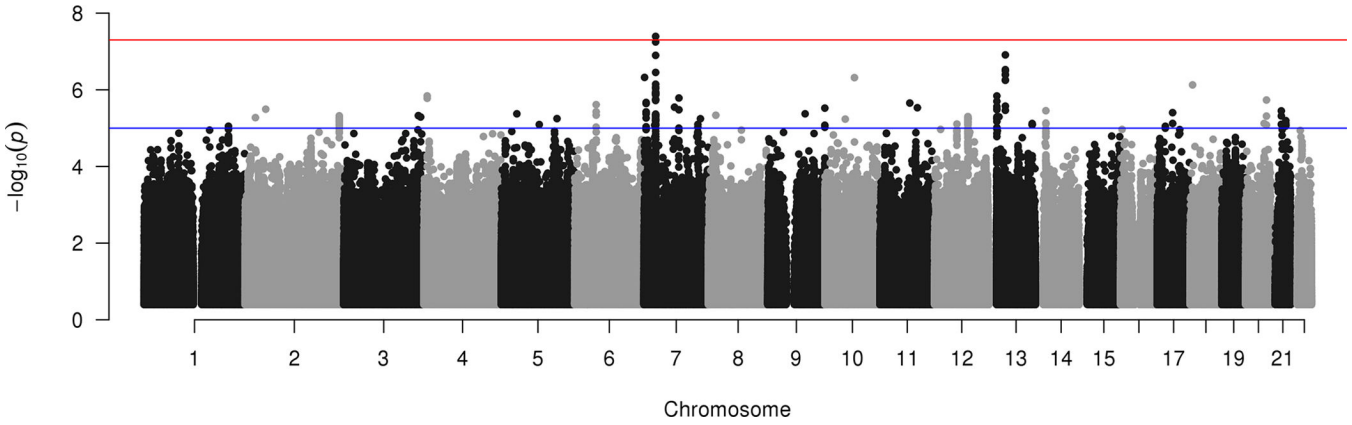

Figure 3. Manhattan plot of $-\log_{10}$ p-values for each association of a genetic variant with rotator cuff disease surgery in the UK Biobank by chromosome location

The red horizontal line indicates the genome-wide significance threshold of $P=5\times10^{-8}$. The blue line represents a threshold of $P=1\times10^{-5}$.

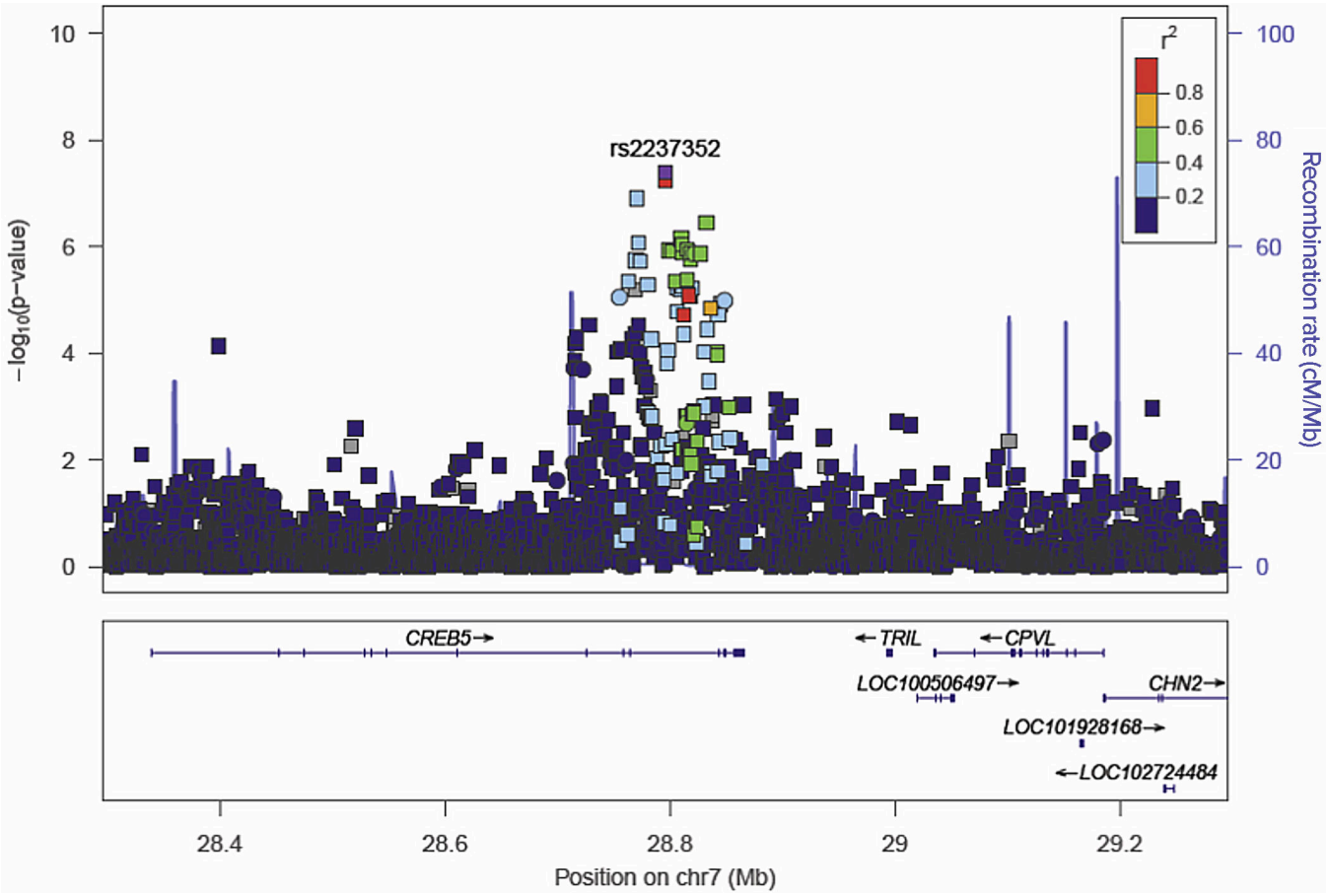

Figure 4. The association for Rotator Cuff Disease (RCD) on chromosome 7.

LocusZoom plot shows the association (left y-axis; log10-transformed p-values) with RCD. Genotyped SNPs are depicted by circles and imputed SNPs are depicted by squares. Shading of the points represent the linkage disequilibrium (r^2, based on the 1000 Genomes Project Europeans) between each SNP and the top SNP, indicated by purple shading. Grey points in the plot represent the lack of LD information between the index SNP (rs2237352) the plotted SNP.

TABLE I

Characteristics of Population of RCD Surgery Cases and Selected Controls from the UK Biobank

	No (%)*	
	Cases	**Controls**
Total	2,917	14,158
Age *(yr)*		
At enrollment	61 (55, 65)	61 (55, 65)
At diagnosis	65 (59, 69)	Not applic.
Sex		
Male	1,503 (51.5%)	7,284 (51.4%)
Female	1,414 (48.5%)	6,874 (48.6%)
Self-reported race/ethnicity group		
White	6 (0.2%)	15 (0.1%)
British	2,747 (94.2%)	13,269 (93.7%)
Irish	79 (2.7%)	393 (2.8%)
Any other White background	85 (2.9%)	481 (3.4%)

*Except for age, which is given as the median (IQR).

Table 2:

Single nucleotide polymorphism (SNP) associations with rotator cuff disease surgery in the UK Biobank

						Cases and Matched Controls					Cases and Expanded Controls		
							Standard Logistic Regression		Conditional Logistic Regression				
Chromosome	**SNP**	**Gene**	**A1**	**A2**	**INFO score**	**A1 frequency**	**OR (95%CI)**	**P-value**	**OR (95%CI)**	**P-value**	**A1 frequency**	**OR (95%CI)**	**P-value**
7	rs2237352	CREB5	C	T	0.9847	0.468	1.17 (1.11, 1.24)	4.04E-08	1.17 (1.10, 1.24)	8.75E-8	0.465	1.16 (1.10, 1.22)	2.29E-08
7	rs12700903	CREB5	G	C	0.9744	0.469	1.17 (1.11, 1.24)	5.63E-08	1.17 (1.10, 1.24)	1.14E-7	0.466	1.16 (1.10, 1.22)	3.69E-08
7	**rs4722837**	**CREB5**	**G**	**A**	**0.9900**	**0.454**	**0.86 (0.81, 0.91)**	**1.26E-07**	**0.86 (0.81, 0.91)**	**1.72E-7**	**0.456**	**0.87 (0.83, 0.92)**	**2.72E-07**
7	rs66539057	CREB5	T	C	1	0.308	1.16 (1.09, 1.23)	1.29E-06	1.16 (1.09, 1.23)	1.35E-06	0.307	1.14 (1.08, 1.20)	5.08E-06

All regression analyses modeled effect of genotype dosage on RCD. Standard logistic regression included covariates for age, sex, follow-up time, and the first 10 population principal components (2,917 cases, 14,158 unique controls). Conditional logistic regression incorporated individuals who could be selected multiple times as controls and conditioned on matched sets of controls with each case to account for all matched covariates (2,917 cases, 14,547 controls). Logistic regression with expanded controls used as controls all non-cases in the UK Biobank that met quality control standards and included covariates for age, sex, follow-up time, and the first 10 population principal components (2,917 cases, 375,560 unique controls).

SNP=single nucleotide polymorphism, MAF=minor allele frequency, OR=odds ratio, A1=Coded allele in regression. A2=Reference allele in regression.

Table 3:

Replication Testing in the UK Biobank of SNPs Associated with Rotator Cuff Disease in the Prior Literature

Chromosome	SNP	Gene	Prior study information			A1	Results from UK Biobank	
			First Author	Cases, N	CG vs. GWAS		OR (95% CI)	p-value
1	rs4654760	ALPL	Peach	22	CG	T	1.00 (0.890, 1.12)	0.958
5	rs3045	ANKH	Peach	22	CG	C	1.01 (0.916, 1.10)	0.910
5	rs1011814	FGF10	Motta	203	CG	T	1.03 (0.967, 1.09)	0.393
5	rs11750845	FGF10	Motta	203	CG	C	1.01 (0.956, 1.07)	0.687
6	rs12527089	SASH1	Tashjian	311	GWAS	T	0.959 (0.836, 1.10)	0.545
8	rs13317	FGFR1	Motta	203	CG	C	0.961 (0.899, 1.03)	0.237
8	rs1800972	DEFB1	Motta	203	CG	C	1.02 (0.956, 1.09)	0.512
9	**rs1590**	**TGFBR1**	**Figueiredo**	**211**	**CG**	**G**	**1.00 (0.940, 1.07)**	**0.937**
9	rs10759753	TNC	Kluger	155	CG	G	0.989 (0.927, 1.05)	0.725
9	rs1138545	TNC	Kluger; Kluger cuff healing	155; 120 (cuff healing study)	CG	T	1.06 (0.982, 1.15)	0.136
9	rs2104772	TNC	Kluger cuff healing, **Figueiredo**	120, **211 (Figueiredo)**	CG	A	1.05 (0.991, 1.11)	0.100
9	rs3789870	TNC	Kluger	155	CG	A	0.994 (0.932, 1.06)	0.865
9	rs7021589	TNC	Kluger	155	CG	C	1.06 (0.984, 1.15)	0.122
9	rs7035322	TNC	Kluger	155	CG	A	1.02 (0.959, 1.09)	0.500
9	rs72758637	TNC	Kluger	155	CG	G	1.07 (0.994, 1.16)	0.070
9	**rs3196378**	**Col5A1**	**Figueiredo**	**211**	**CG**	**C**	**1.02 (0.967, 1.08)**	**0.424**
11	rs12574452	FGF3	Motta	203	CG	A	1.04 (0.982, 1.11)	0.174
11	rs1799750	MMP1	Assuncao	64	CG	Not measured or did not meet quality control filtering criteria in UK Biobank		
11	rs3025058	MMP3	Assuncao	64	CG			
11	**rs679620**	**MMP3**	**Figueiredo**	**211**	**CG**	**C**	**1.01 (0.957, 1.07)**	**0.658**
14	rs10132091	ESRRB	Bonato	49	CG	C	0.991 (0.936, 1.05)	0.760
14	rs1676303	ESRRB	Motta, Bonato	203 (Motta); 16 (Bonato)	CG	C	1.00 (0.915, 1.10)	0.964
14	rs17583842	ESRRB	Teerlink, Tashjian cuff healing	175 (Teerlink); 30 (Tashjian cuff healing)	CG	C	0.940 (0.879, 1.01)	0.072

Chromosome	SNP	Gene	Prior study information			A1	Results from UK Biobank	
			First Author	Cases, N	CG vs. GWAS		OR (95% CI)	p-value
14	rs4903399	ESRRB	Motta, Bonato	203 (Motta); 49 (Bonato)	CG	T	1.03 (0.959, 1.11)	0.418
16	**rs2285053**	**MMP2**	**Figueiredo**	**211**	**CG**	**T**	**1.08 (0.985, 1.19)**	**0.101**
16	rs71404070		Roos	8357	GWAS	A	0.993 (0.866, 1.14)	0.919
17	rs820218	SAP30BP	Tashjian	311	GWAS	A	0.934 (0.879, 0.991)	0.025
17	**rs2277698**	**TIMP2**	**Figueiredo**	**211**	**CG**	**T**	**0.870 (0.795, 0.952)**	**0.002**
19	**rs1800470**	**TFGB1**	**Figueiredo**	**211**	**CG**	**G**	**1.02 (0.961, 1.08)**	**0.532**
19	**rs1800469**	**TFGB1**	**Figueiredo**	**211**	**CG**	**A**	**1.01 (0.951, 1.08)**	**0.710**
20	**rs17576**	**MMP9**	**Figueiredo**	**211**	**CG**	**G**	**1.01 (0.949, 1.07)**	**0.821**

CG=candidate gene study, GWAS=genome-wide association study