Identification of a novel genetic marker for risk of degenerative rotator cuff disease surgery in the UK Biobank

Elizabeth L Yanik  
Washington University School of Medicine in St. Louis

Jay D Keener  
Washington University School of Medicine in St. Louis

Shiow J Lin  
Washington University School of Medicine in St. Louis

Graham A Colditz  
Washington University School of Medicine in St. Louis

Rick W Wright  
Washington University School of Medicine in St. Louis

See next page for additional authors

Follow this and additional works at: https://digitalcommons.wustl.edu/open_access_pubs

Please let us know how this document benefits you.

Recommended Citation
https://digitalcommons.wustl.edu/open_access_pubs/10424

This Open Access Publication is brought to you for free and open access by Digital Commons@Becker. It has been accepted for inclusion in Open Access Publications by an authorized administrator of Digital Commons@Becker. For more information, please contact vanam@wustl.edu.
Authors
Elizabeth L Yanik, Jay D Keener, Shiow J Lin, Graham A Colditz, Rick W Wright, Bradley A Evanoff, Nitin B Jain, and Nancy L Saccone
Identification of a Novel Genetic Marker for Risk of Degenerative Rotator Cuff Disease Surgery in the UK Biobank

Elizabeth L. Yanik, PhD, ScM, Jay D. Keener, MD, Shio J. Lin, MS, Graham A. Colditz, DrPH, MD, Rick W. Wright, MD, Bradley A. Evanoff, MD, MPH, Nitin B. Jain, MD, MSPH, and Nancy L. Saccone, PhD

Investigation performed at Washington University School of Medicine, St. Louis, Missouri

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

Methods: Cases with surgery for degenerative RCD were identified using linked hospital records. The cases were defined as an International Classification of Diseases, Tenth Revision (ICD-10) code of M75.1 determined by a trauma/orthopaedic specialist and surgery consistent with RCD treatment. Cases were excluded if a diagnosis of traumatic injury had been made during the same hospital visit. For each case, up to 5 controls matched by age, sex, and follow-up time were chosen from the UK Biobank. Analyses were limited to European ancestry individuals who were not third-degree or closer relations. We used logistic regression to test for genetic association of 674,405 typed and >10 million imputed markers, after 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 1 genome-wide significant signal (p < 5 × 10^{-8}) for a novel locus tagged by rs2237352 in the CREB5 gene on chromosome 7 (odds ratio [OR] = 1.17, 95% confidence interval [CI] = 1.11 to 1.24). The 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 sample to include additional unmatched non-cases, rs2237352 and another SNP in 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 an important next step.

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

Level of Evidence: Prognostic Level III. See Instructions for Authors for a complete description of levels of evidence.

Rotator cuff disease (RCD) is the most common cause of shoulder disability[^1^], but studies identifying risk factors for symptomatic RCD have been scarce. Prevention strategies aimed at high-risk groups could dramatically impact disease burden. Accumulating evidence indicates that familial predisposition influences the risk of developing degenerative RCD[^2^][^3^]. However, knowledge of specific RCD genetic markers is limited. Investigators have evaluated candidate genes and discovered associations with this condition in previous studies.
The first genome-wide association study (GWAS) for RCD, of which we are aware, detected 2 associated single-nucleotide polymorphisms (SNPs) involved in apoptosis, but this study of <350 patients with RCD had limited statistical power. The authors of a second GWAS were unable to replicate the associations found in the first, and they identified a new SNP associated with RCD. While this study was much larger (8,357 rotator cuff injuries), the definition for RCD was nonspecific, with use of International Classification of Diseases (ICD) codes that might capture shoulder pain from other diseases.

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.

Materials and Methods

Our study population was derived from the UK Biobank, a population-based prospective cohort of approximately 500,000 U.K. residents. Participants 40 to 69 years of age were recruited nationwide from 2006 to 2010 through invitations mailed to people registered with the National Health Service (NHS). At enrollment, participants gave informed consent and whole blood samples were collected. NHS hospital records were linked to the UK Biobank, providing information on inpatient diagnoses and procedures from 2006 to 2017. Diagnoses were coded using the ICD, Tenth Revision (ICD-10). Procedures were coded using the Office of Population Censuses and Surveys Classification, Fourth 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. 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 the 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.

We selected cases and controls from the remaining group of 487,324 people. As >90% of participants self-reported European/White ancestry, the control population was limited to that group to reduce population stratification. Of the 10 population principal components, the first 10 population principal components were graphed, and a genomic inflation factor was calculated to check for bias. We required a standard genome-wide significance threshold of $5 \times 10^{-8}$. 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. For these markers, a Bonferroni-adjusted $p$ value of <0.05 and an odds ratio (OR) indicating an association in the same direction as the original finding was necessary.
were made after a median of 5 years (interquartile range [IQR] = 3 to 6 years, range = 0 to 10 years) of follow-up (i.e., after enrollment in the UK Biobank). The median age at diagnosis was 65 years old (IQR = 59 to 69 years, range = 41 to 78 years), and 48.5% of the cases were women (Table 1). Cases were matched with a 1:5 ratio to 14,158 unique controls on the basis of follow-up, age, and sex. For conditional logistic regression to represent incidence-density sampling, some individuals could be selected as controls multiple times, resulting in 14,547 controls.

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

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

We observed 1 novel genome-wide significant signal (p = 4.04 × 10⁻⁷) at SNP rs2237352 in the CREBS5 gene on chromosome 7 (OR = 1.17, 95% confidence interval [CI] = 1.11 to 1.24; Table II, Fig. 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 SNP rs12700903 in the CREBB5 gene (OR = 1.17, 95% CI = 1.11 to 1.24, p = 5.63 × 10⁻⁵), which is in strong linkage disequilibrium with rs237352 (r² = 0.98). Thus, both SNPs represent the same statistical signal. The most significant directly assayed SNP in the CREBS5 gene was rs66539057, but the association was not significant at a genome-wide level (OR = 1.16, 95% CI = 1.09 to 1.23, p = 1.29 × 10⁻⁴). Figure 4 shows a detailed view of the associated

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

<table>
<thead>
<tr>
<th></th>
<th>No (%)*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cases</td>
</tr>
<tr>
<td>Total</td>
<td>2,917</td>
</tr>
<tr>
<td>Age (yr)</td>
<td></td>
</tr>
<tr>
<td>At enrollment</td>
<td>61 (55, 65)</td>
</tr>
<tr>
<td>At diagnosis</td>
<td>65 (59, 69)</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>1,503 (51.5%)</td>
</tr>
<tr>
<td>Female</td>
<td>1,414 (48.5%)</td>
</tr>
</tbody>
</table>
region, with the lead SNP rs2237352 having strong to moderate linkage disequilibrium with additional SNPs in the region. The strongest signal for an SNP having modest linkage disequilibrium with rs2237352 was for rs4722837 (OR = 0.86, p = 1.26 \times 10^{-7}, r^2 = 0.38). Results were similar in conditional logistic regression (rs2237352: OR = 1.17, 95% CI = 1.10 to 1.24; rs12700903: OR = 1.17, 95% CI = 1.10 to 1.24) (Table II).

After analyses with adjustment for the rs2237352 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.

Fig. 2
Q-Q plot comparing the observed p values (pval) with the expected distribution of p values for each association of a genetic variant with RCD 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.

study. After Bonferroni adjustment for 29 replication attempts, neither SNP remained significant.

After we expanded the control group to include the larger, unmatched cohort of non-cases (n = 375,560), rs2237352 remained genome-wide significant (p = 2.29 \times 10^{-8}, Table II). Additionally, the association of rs12700903 with surgery for degenerative RCD became genome-wide significant (p = 3.69 \times 10^{-8}).

There were 735 patients with RCD surgery who were ≤60
### TABLE II SNP Associations with RCD Surgery in the UK Biobank

<table>
<thead>
<tr>
<th>Chromosome</th>
<th>SNP</th>
<th>Gene</th>
<th>A1</th>
<th>A2</th>
<th>A1 Score</th>
<th>A1 Frequency</th>
<th>OR (95% CI)</th>
<th>P Value (Conditional Logistic Regression)</th>
<th>OR (95% CI)</th>
<th>P Value (Standard Logistic Regression)</th>
<th>Cases and Expanded Controls</th>
<th>Cases and Matched Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>rs2237352</td>
<td>CREB5</td>
<td>C</td>
<td>T</td>
<td>0.9567</td>
<td>0.468</td>
<td>1.17 (1.11, 1.24)</td>
<td>8.75 x 10^-8</td>
<td>1.17 (1.10, 1.24)</td>
<td>8.75 x 10^-8</td>
<td>0.465</td>
<td>1.16 (1.10, 1.22)</td>
</tr>
<tr>
<td>7</td>
<td>rs12700903</td>
<td>CREB5</td>
<td>G</td>
<td>C</td>
<td>0.9744</td>
<td>0.469</td>
<td>1.17 (1.11, 1.24)</td>
<td>5.63 x 10^-8</td>
<td>1.17 (1.10, 1.24)</td>
<td>1.14 x 10^-7</td>
<td>0.466</td>
<td>1.16 (1.10, 1.22)</td>
</tr>
<tr>
<td>7</td>
<td>rs4722837</td>
<td>CREB5</td>
<td>G</td>
<td>A</td>
<td>0.9900</td>
<td>0.454</td>
<td>0.86 (0.81, 0.91)</td>
<td>1.26 x 10^-7</td>
<td>0.86 (0.81, 0.91)</td>
<td>1.72 x 10^-7</td>
<td>0.456</td>
<td>0.87 (0.83, 0.92)</td>
</tr>
<tr>
<td>7</td>
<td>rs66539057</td>
<td>CREB5</td>
<td>T</td>
<td>C</td>
<td>0.308</td>
<td>1.16 (1.09, 1.23)</td>
<td>1.29 x 10^-6</td>
<td>1.16 (1.09, 1.23)</td>
<td>1.35 x 10^-6</td>
<td>0.307</td>
<td>1.14 (1.06, 1.20)</td>
<td>5.08 x 10^-6</td>
</tr>
</tbody>
</table>

*All regression analyses modeled the 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 and 14,158 unique controls). Conditional logistic regression incorporated individuals who could be selected multiple times as controls and was conditioned on matched sets of controls with each case to account for all matched covariates (2,917 cases and 14,158 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 and 375,560 unique controls). A1 = coded allele in regression, and A2 = reference allele in regression.

A novel association between the SNP rs2237352 and surgery for degenerative RCD. The SNP rs12700903 was also associated with surgery for degenerative RCD after expansion of our control group, and it represents the same signal as SNP rs2237352. Both SNPs are located in the CREB5 gene, which encodes a protein that is part of the CAMP response element-binding protein family. CREB5 is a transcription factor involved in cell growth, proliferation, and differentiation. CREB5 expression has been associated with plasma interleukin-6 levels and may influence inflammatory response genes. As CREB family proteins influence expression of other genes, there may be numerous genetic mutations that could influence the same biological pathways. Further research confirms this association, one would expect the genetic risk for RCD to be highly polygenic as is common for most complex traits. Differential CREB expression has also been specifically documented in fibroblasts, lending further evidence that mutations in this gene could be of importance for tendon injury and repair.

After adjustment for rs2237352 in our models, no additional signals were detected, which is consistent with this region harboring 1 primary locus associated with degenerative RCD. CREB5 SNPs in weaker linkage disequilibrium 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. 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. To our knowledge, 2 GWAS RCD studies in independent populations have been conducted. 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 29 SNPs
from prior studies (Table III) that we could evaluate in the UK Biobank, only 1 (rs820218) demonstrated an association in the same direction as it did in the prior study\textsuperscript{19}, with a nominal uncorrected \( p \) value of \(< 0.05\), while none reached genome-wide significance. SNP rs820218 is located in the SAP30BP gene, which encodes a transcriptional regulator protein involved in cell death and apoptosis\textsuperscript{34,35}. Numerous studies have shown increased tendon cell apoptosis related to rotator cuff tearing\textsuperscript{36-38}. However, as only 1 SNP out of 29 demonstrated a
<table>
<thead>
<tr>
<th>Chromosome</th>
<th>SNP</th>
<th>Gene</th>
<th>Prior Study Information</th>
<th>Results from UK Biobank</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>First Author</td>
<td>No. of Cases</td>
</tr>
<tr>
<td>1</td>
<td>rs4654760</td>
<td>ALPL</td>
<td>Peach\textsuperscript{26}</td>
<td>22</td>
</tr>
<tr>
<td>5</td>
<td>rs3045</td>
<td>ANKH</td>
<td>Peach\textsuperscript{26}</td>
<td>22</td>
</tr>
<tr>
<td>5</td>
<td>rs1011814</td>
<td>FGFr10</td>
<td>Motta\textsuperscript{11}</td>
<td>203</td>
</tr>
<tr>
<td>5</td>
<td>rs11750845</td>
<td>FGFr10</td>
<td>Motta\textsuperscript{11}</td>
<td>203</td>
</tr>
<tr>
<td>6</td>
<td>rs12527089</td>
<td>SASH1</td>
<td>Tashjian\textsuperscript{14}</td>
<td>311</td>
</tr>
<tr>
<td>8</td>
<td>rs13317</td>
<td>FGFr1</td>
<td>Motta\textsuperscript{11}</td>
<td>203</td>
</tr>
<tr>
<td>8</td>
<td>rs1800972</td>
<td>DEF1</td>
<td>Motta\textsuperscript{11}</td>
<td>203</td>
</tr>
<tr>
<td>9</td>
<td>rs1599</td>
<td>TGFBR1</td>
<td>Figueiredo\textsuperscript{27}</td>
<td>211</td>
</tr>
<tr>
<td>9</td>
<td>rs10759753</td>
<td>TNC</td>
<td>Kluger\textsuperscript{13}</td>
<td>155</td>
</tr>
<tr>
<td>9</td>
<td>rs1138545</td>
<td>TNC</td>
<td>Kluger\textsuperscript{13,28}</td>
<td>155\textsuperscript{13,28}</td>
</tr>
<tr>
<td>9</td>
<td>rs2104772</td>
<td>TNC</td>
<td>Kluger\textsuperscript{28}, Figueiredo\textsuperscript{27}</td>
<td>120\textsuperscript{28}, 211\textsuperscript{27}</td>
</tr>
<tr>
<td>9</td>
<td>rs3789870</td>
<td>TNC</td>
<td>Kluger\textsuperscript{13}</td>
<td>155</td>
</tr>
<tr>
<td>9</td>
<td>rs7021589</td>
<td>TNC</td>
<td>Kluger\textsuperscript{13}</td>
<td>155</td>
</tr>
<tr>
<td>9</td>
<td>rs7035322</td>
<td>TNC</td>
<td>Kluger\textsuperscript{13}</td>
<td>155</td>
</tr>
<tr>
<td>9</td>
<td>rs72758637</td>
<td>TNC</td>
<td>Kluger\textsuperscript{13}</td>
<td>155</td>
</tr>
<tr>
<td>9</td>
<td>rs3196378</td>
<td>Col5A1</td>
<td>Figueiredo\textsuperscript{27}</td>
<td>211</td>
</tr>
<tr>
<td>11</td>
<td>rs12574452</td>
<td>FGF3</td>
<td>Motta\textsuperscript{11}</td>
<td>203</td>
</tr>
<tr>
<td>11</td>
<td>rs1799750</td>
<td>MMP3</td>
<td>Assunção\textsuperscript{56}</td>
<td>64</td>
</tr>
<tr>
<td>11</td>
<td>rs3025058</td>
<td>MMP3</td>
<td>Assunção\textsuperscript{56}</td>
<td>64</td>
</tr>
<tr>
<td>11</td>
<td>rs679620</td>
<td>MMP3</td>
<td>Figueiredo\textsuperscript{27}</td>
<td>211</td>
</tr>
<tr>
<td>14</td>
<td>rs10132091</td>
<td>ESRRB</td>
<td>Bonato\textsuperscript{12}</td>
<td>49</td>
</tr>
<tr>
<td>14</td>
<td>rs1676503</td>
<td>ESRRB</td>
<td>Motta\textsuperscript{11,12}</td>
<td>203\textsuperscript{11,12}</td>
</tr>
<tr>
<td>14</td>
<td>rs17583842</td>
<td>ESRRB</td>
<td>Teenlink\textsuperscript{29}, Tashjian\textsuperscript{30}</td>
<td>175\textsuperscript{29}, 30\textsuperscript{30}</td>
</tr>
<tr>
<td>14</td>
<td>rs4903399</td>
<td>ESRRB</td>
<td>Motta\textsuperscript{11,12}</td>
<td>203\textsuperscript{11,12}</td>
</tr>
<tr>
<td>16</td>
<td>rs2285053</td>
<td>FGF2</td>
<td>Figueiredo\textsuperscript{27}</td>
<td>211</td>
</tr>
<tr>
<td>16</td>
<td>rs71400470</td>
<td>Reo\textsuperscript{15}</td>
<td>8,357</td>
<td>GWAS</td>
</tr>
<tr>
<td>17</td>
<td>rs820218</td>
<td>SAP30BP</td>
<td>Tashjian\textsuperscript{14}</td>
<td>311</td>
</tr>
<tr>
<td>17</td>
<td>rs2277698</td>
<td>TIMP2</td>
<td>Figueiredo\textsuperscript{27}</td>
<td>211</td>
</tr>
<tr>
<td>19</td>
<td>rs1800470</td>
<td>TFGB1</td>
<td>Figueiredo\textsuperscript{27}</td>
<td>211</td>
</tr>
<tr>
<td>19</td>
<td>rs1800469</td>
<td>TFGB1</td>
<td>Figueiredo\textsuperscript{27}</td>
<td>211</td>
</tr>
<tr>
<td>20</td>
<td>rs175576</td>
<td>MMP9</td>
<td>Figueiredo\textsuperscript{27}</td>
<td>211</td>
</tr>
</tbody>
</table>

*CG = candidate gene study, and GWAS = genome-wide association study. †Not measured or did not meet quality-control filtering criteria in the UK Biobank.

There are important limitations to the current study. First, the UK Biobank contains predominantly white participants, limiting the generalizability of findings to other populations. Second, the study does not include any outcomes and thus could not evaluate how genetics influenced treatment outcomes.
than those in other large RCD GWASs, and we did not appear to be substantially undercapturing RCD surgery cases based on rates in the literature. 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 the UK Biobank will improve in the future as plans are in place for whole genome sequencing of the population.

A more comprehensive understanding of genetic susceptibility to 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 an 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 CREBS gene associated with surgery for degenerative RCD in a general population sample of the U.K. 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 the UK Biobank could allow future evaluation of risk models incorporating genetics, non-genetic characteristics, and gene-environment interactions. 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.

Elizabeth L. Yanik, PhD, ScM
Jay D. Keener, MD
Shiou J. Lin, MS
Graham A. Colditz, DrPH, MD
Rick W. Wright, MD
Bradley A. Evanoff, MD, MPH
Nitin B. Jain, MD, MSc
Nancy L. Saccoone, PhD

1Departments of Orthopaedic Surgery (E.L.Y., J.D.K., and R.W.W.), Surgery (E.L.Y. and G.A.C.), Genetics (S.L.L. and N.L.S.), and General Medical Studies (B.A.E.), Washington University School of Medicine, St. Louis, Missouri
2Department of Physical Medicine and Rehabilitation, University of Texas Southwestern, Dallas, Texas

Email address for E.L. Yanik: yanike@wustl.edu

ORCID iD for E.L. Yanik: 0000-0002-5835-0201
ORCID iD for J.D. Keener: 0000-0002-1665-4346
ORCID iD for S.J. Lin: 0000-0002-4292-1751
ORCID iD for G.A. Colditz: 0000-0002-7307-0291
ORCID iD for R.W. Wright: 0000-0003-4018-7132
ORCID iD for B.A. Evanoff: 0000-0003-085X-333X
ORCID iD for N.B. Jain: 0000-0002-4362-8522
ORCID iD for N.L. Saccoone: 0000-0002-4710-3926

References


