

Regulatory VCAN Polymorphism is Associated with Shoulder Pain and Disability in Breast Cancer Survivors

Trevor Shepherd Mafu

University of Cape Town <https://orcid.org/0000-0002-5485-5747>

Alison V September

University of Cape Town

Delva Shamley (✉ delva.shamley@uct.ac.za)

University of Cape Town <https://orcid.org/0000-0003-2355-6629>

Primary research

Keywords: proteoglycans, variant, connective tissue, chronic pain, shoulder dysfunction

Posted Date: February 24th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-233765/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background and Purpose

Shoulder morbidity following breast cancer treatment is multifactorial. Despite several treatment- and patient-related factors implicated, unexplained inter-individual variability exists in the development of such morbidity. Given the paucity of relevant genetic studies, we investigate the role of polymorphisms in candidate proteoglycan genes.

Patients and methods

We conducted a cross-sectional study on 254 South African breast cancer survivors, to evaluate associations between shoulder pain/disability and ten single nucleotide polymorphisms (SNPs) within four proteoglycan genes: *ACAN* (rs1126823 G>A, rs1516797 G>T, rs2882676 A>C); *BGN* (rs1042103 G>A, rs743641 A>T, rs743642 G>T); *DCN* rs516115 C>T; and *VCAN* (rs11726 A>G, rs2287926 G>A, rs309559). Participants were grouped into no–low and moderate–high shoulder pain/disability based on total pain/disability scores: <30 and \geq 30, respectively using the shoulder pain and disability index (SPADI).

Results

The GG genotype of *VCAN* rs11726 was independently associated with an increased risk of being in the moderate to high shoulder pain ($P=0.005$, OR=2.326, 95% CI=1.259 - 4.348) or disability ($P=0.011$, OR=2.439, 95% CI=1.235 - 4.762) categories, after adjusting for participants' age. In addition, the T-T-G inferred allele combination of *BGN* (rs74364 - rs743642) - *VCAN* rs11726 was associated with an increased risk of being in the moderate to high shoulder disability category ($P=0.002$, OR=2.347, 95% CI=1.215 - 4.534).

Conclusion

Our study is first to report that *VCAN* rs11726, independently or interacting with *BGN* polymorphisms, is associated with shoulder pain or disability in breast cancer survivors. Whereas our findings suggest an involvement of proteoglycans in the etiology of shoulder pain/disability, further studies are recommended.

Introduction

Shoulder pain and disability are common chronic upper-limb morbidities among female breast cancer survivors [1, 2]. We recently reported that 74 % and 62 % of breast cancer survivors report some level of pain and functional disability, respectively, at least 1 year post-surgery [1]. The most recent systematic review of upper-limb morbidity amongst breast cancer survivors by Hidding *et al.* [2] estimates prevalence of upper-limb pain and disability (reduced range of motion) at 9 – 68 % and 6 – 31 %, respectively, beyond 1 year post-surgery. Shoulder pain and disability are strongly correlated [1], and may persist

beyond 7 years following treatment [3]. Given that such morbidities have a negative impact on the quality of life of affected individuals [4], understanding the underlying etiology remains an urgent need.

The etiology underlying shoulder pain/disability remains poorly understood. Contributing treatment related factors include: type of breast surgery, type of axillary surgery and adjuvant therapy, with absolute risk increases of 1 – 21 % for persistent pain [2, 5, 6]; Patient related factors include: age, presence of pre-operative pain, presence of acute post-operative pain and genetic predisposition, with absolute risk increases of 2 – 7 % for persistent pain [6, 7]. Nonetheless, there remains a paucity of studies investigating the role of genetic factors in the inter-individual variability in developing shoulder pain/disability amongst breast cancer survivors.

A growing body of evidence, from non-cancer-related conditions, supports the role of polymorphisms within genes encoding structural and regulatory extracellular matrix (ECM) proteins in modulating susceptibility to shoulder pathology [8-11]. Although no associations between polymorphisms in proteoglycan encoding genes and non-cancer-related shoulder conditions have been reported to date, proteoglycans are important ECM components whose expression levels are altered in such conditions [12, 13]. In particular, changes in expression of the proteoglycans: aggrecan (ACAN), versican (VCAN), biglycan (BGN) and decorin (DCN), have been reported in rotator cuff disease [12, 13]. Moreover, associations have been reported between proteoglycan gene polymorphisms and other connective tissue conditions such as anterior cruciate ligament ruptures [10, 14-16]. The hylectan proteoglycans, including ACAN and VCAN are important structural components in connective tissues such as tendons and ligaments. The small leucine-rich proteoglycans (SLRPs) including BGN and DCN regulate collagen fibrillogenesis, and are important modulators of the angiogenesis associated and the transforming growth factor beta (TGF- β) signaling pathways amongst others [17]. Given the role of proteoglycans in the ECM of connective tissues, functional polymorphisms in proteoglycan encoding genes may, perhaps, lead to altered signaling and/or biomechanical properties in tissues such as tendons or ligaments of the shoulder. Indeed, studies on animal models have demonstrated that changes in expression of decorin and biglycan alter mechanical properties of tendons including failure load, stiffness, dynamic modulus and viscosity [18]. We hypothesize that polymorphisms in proteoglycan encoding genes may be associated with shoulder pain/disability amongst breast cancer survivors. Our aim, therefore, was to investigate the association between candidate gene polymorphisms within proteoglycan encoding genes and shoulder pain/disability following breast cancer treatment in women.

Methods

Study design

Pilot, cross-sectional, genetic association study based on the candidate gene approach.

Participants and Setting

A total of 254 participants were conveniently recruited in the year period 2013 – 2018, from the waiting room of the Oncology Clinic of a tertiary public teaching hospital in South Africa. All eligible participants (Table 1) agreeing to participate gave written informed consent. The recruited participants self-identified as ‘mixed-ancestry’ ethnicity, a rich genetic admixture ancestrally derived from immigrants from Western Europe, West Africa, Asia and the indigenous Southern African populations [19].

Study procedures

This study received ethical approval from the Human Research Ethics Committee at the University of Cape Town (HREC REF: 359/2019). Study procedures have been previously reported [7]. Briefly, eligible consented participants completed the Shoulder Pain and Disability Index (SPADI) questionnaire and had their bloods drawn by venipuncture at the cubital fossa of the unaffected side using EDTA vacutainer tubes. Whole blood samples were immediately stored at -20°C until total DNA extraction using the method described by Lahiri *et al* [20]. Extracted DNA was stored long-term at -20°C. Relevant information for each participant including age, tumor grade, surgery data, and adjuvant therapy data were obtained from participants’ medical records.

Patient reported outcome measure

The primary outcome measure in this study was the SPADI, a validated and reliable patient reported questionnaire with two domains: Pain (5 items) and Disability (8 items) [21, 22]. Participants rated pain or difficulty associated with specific activities of daily living on a visual analog scale (VAS) of 0 (no pain/difficulty) – 10 (extreme pain/difficulty). Symptom scores for both SPADI domains were reported as percentages of possible total scores [22].

Pain and disability scores were categorized according to score effects on activities of daily living and clinical relevance [7]. The reference ‘no – low’ category consisted of participants with SPADI pain/disability scores <30 whereas the case ‘moderate – high’ category consisted of participants with SPADI pain/disability scores ≥ 30 .

Genetic variables

Exposures in this study were the total genotypes obtained from genotyping single nucleotide polymorphisms (SNPs) within four candidate proteoglycan genes: *ACAN* (rs1126823 G>A, rs1516797 G>T, rs2882676 A>C); *BGN* (rs1042103 G>A, rs743641 A>T, rs743642 G>T); *DCN* (rs516115 C>T); and *VCAN* (rs11726 A>G, rs2287926 G>A, rs309559).

Single Nucleotide Polymorphism (SNP) selection

SNPs with global minor allele frequency >0.15 in the ENSEMBL database (<http://www.ensembl.org>) were selected for investigation based on meeting one or more of the following criteria:

- Identified from a whole exome sequencing project on risk factors for tendinopathy or musculoskeletal soft tissue injuries [23].
- Functional significance, based on reported effects on gene expression or protein function.
- Located in regulatory gene regions.
- Previous associations with multifactorial soft-tissue shoulder conditions.

A total of ten SNPs within four proteoglycan encoding genes were included (Tables 4 and 5). In order to ensure robust genetic association analyses, only SNP call rates of >95% and Hardy-Weinberg p-values >0.05 were included.

Genotype determination

Genotyping was performed using TaqMan™ assays (Applied Biosystems) in 96-well plates, following manufacturer's instructions in a QuantStudio™ 3 Real-Time PCR System (Applied Biosystems) at the Division of Exercise Science and Sports Medicine, University of Cape Town. Both negative controls (no DNA sample), positive controls (DNA of known genotypes) and replicates (sample duplicates) were included in every plate to evaluate the reliability of the PCR and detect potential genotyping errors. The genotyping data were analyzed on Thermo Fisher Cloud genotyping analysis Software Version: 3.3.0-SR2-build 21 with automatic genotype calling for the 9 SNPs: *ACAN* (rs1126823 G>A, rs156797 G>T); *BGN* (rs1042103 G>A, rs743641 A>T, rs743642 G>T), *DCN* (rs516115 C>T) and *VCAN* (rs11726 A>G, rs2287926 G>A, rs309559). Due to less efficient amplification for the *ACAN* rs2882676 A>C SNP, genotypes were manually called and compared with the manual calls of an independent blinded technical support member with 99.7% similarity.

Bias

Nine percent (23, out of 254) of participants could not provide bloods because they were lost after consent when they went for further medical examination in the clinic. Although there may be differences between participants who provided bloods and those who did not, it is unlikely as all participants were randomly identified and consented.

Statistical analysis

The calculation of sample size for this study, using QUANTO version 1.2.469 [24], was described previously [7]. A sample size of N=231 was regarded likely sufficient to detect odds ratios of ≥ 2.5 for

allele frequencies ≥ 0.15 , assuming an expected average baseline risk for shoulder pain (32%) and disability (25%), for dominant or additive genetic models [7].

Demographic and clinical data were analyzed using Statistica version 13.2.70 [25]. Mann Whitney U tests were used to evaluate differences in quantitative characteristics between the shoulder pain/disability categories, given that the data was non-parametric. Fisher's exact and Chi-square analyses were performed to evaluate differences in categorical demographic and clinical characteristics between the shoulder pain/disability categories.

The genotype data were analyzed using R Studio version 1.3.895 running R version 3.6.3 [26, 27]. Chi-square and Fisher's exact tests were used to evaluate differences in the genotype, allele and inferred haplotype frequencies between the shoulder pain/disability categories. Hardy-Weinberg equilibrium (HWE) and linkage disequilibrium (LD) were calculated using R package 'genetics' version 1.3.8.1.2 [28]. Logistic regression analyses were performed using R package 'SNPassoc' version 1.9-2 to evaluate the association between SNP genotype and shoulder pain/disability category membership [29]; The best model (with the lowest Akaike Information Criterion (AIC)) was chosen among dominant, recessive and log-additive models. Using the R package 'haplo.stats' Version 1.7.9 [30], inferred haplotypes for the *ACAN*, *BGN* and *VCAN* polymorphisms were constructed using the genotype data for each SNP investigated. To investigate possible gene-gene interactions in modulating risk for shoulder pain/disability, inferred allele combinations were constructed using the relevant genotype data for the genes. The choice of SNPs for inferred allele combination construction was based on stepwise backward elimination logistic regression analysis. In each step, the least informative SNPs whose exclusion lowered, and therefore improved, the AIC of the model were removed until the last three SNPs representing the best model for shoulder pain or disability with three SNPs. To avoid saturating the models while controlling for confounding, only participants' age, which was shown to be associated with our primary outcomes, was included in all multivariate regression models. For all inferred haplotypes or allele combinations, a low haplotype frequency cut-off of 4% was used to improve validity. Stepwise regression analyses were performed using R package 'MASS' version 7.3-51.5 [31]. R package 'ggplot2' version 3.2.1 was used to produce all graphs [32]. The level of significance was set as $p < 0.05$.

Results

Differences in clinical and demographic characteristics between pain/disability categories

Participants in the moderate – high shoulder pain category were significantly younger compared with those in the no – low shoulder pain category 53.8 (45.3 – 64.3) vs. 60.8 (53.5 – 65.5), $p=0.001$ (Table 2). Similarly, participants in the moderate – high shoulder disability category were significantly younger compared with those in the no – low disability category (54.4 (45.0 – 64.9) vs. 60.4 (53.2 – 65.2), $p=0.014$) (Table 3). However, no significant differences ($p > 0.05$) were noted between participants in the no – low and moderate – high shoulder pain/disability categories for all other variables. Despite being

statistically insignificant, a lower proportion of participants in the moderate – high shoulder pain/disability category underwent the more aggressive surgeries: mastectomy and axillary lymph node dissection, compared to those in the no – low shoulder pain/disability category (Tables 2 and 3). In addition, a higher proportion of participants in the moderate – high shoulder pain/disability category received adjuvant chemotherapy compared to those in the no – low shoulder pain/disability category (Tables 2 and 3). Receipt of adjuvant radiotherapy was only notable for shoulder pain categories, with a higher proportion of participants in the moderate – high category receiving the same compared to participants in the no – low category (Table 2).

Interestingly, we noted that ($p=0.014$) a higher proportion of participants with the GG (88.2%, $n=60$) genotype for *ACAN* rs1126823 G>A received hormonal therapy compared with those with AA or AG (73.0%, $n=111$) genotypes (Supplementary Table 1). A significantly ($p=0.034$) lower proportion of participants with the TT (58.3%, $n=14$) genotype for *BGN* rs743641 A>T received hormonal therapy compared with participants with AA or AT (80.1%, $n=157$) genotypes. Furthermore, a significantly ($p=0.001$) higher proportion of participants with the AA (92.9%, $n=65$) genotype of *ACAN* rs2882676 A>C had mastectomy compared with those with AC or CC (73.1%, $n=114$) genotypes. Whereas, individuals with a TT (66.7%, $n=44$) genotype for *DCN* rs516115 C>T were significantly ($p=0.007$) less likely to have mastectomy compared with CT or CC (84.0%, $n=136$) genotype carriers.

Genotype/allele frequency distributions between shoulder pain/disability categories

The genotype frequencies of the *VCAN* rs11726 A>G polymorphism were significantly different ($p<0.05$) between the no – low and moderate – high categories for both shoulder pain and disability, including after adjustment for participants' age (Tables 4 and 5). In particular, the GG genotype of *VCAN* rs11726 A>G was significantly more common ($p=0.005$; OR=2.326, 95% CI= 1.250 – 4.348) in the moderate – high shoulder pain category (48%) in comparison with the no – low shoulder pain category (29%) (Table 4). Similarly, the GG genotype of *VCAN* rs11726 A>G was significantly more common ($p=0.011$; OR= 2.439, 95% CI= 1.235 – 4.762) in the moderate – high shoulder disability category (51%) in comparison with the no – low shoulder disability category (30%) (Table 5). No significant differences were noted in allele frequency distributions between the no – low and moderate – high categories for the *VCAN* rs11726 A>G polymorphism (Tables 4 and 5). However, there was a trend ($p=0.069$) towards over-representation of the A allele of *VCAN* rs11726 A>G in the no – low shoulder disability category (44%) in comparison with the moderate – high disability category (33%) (Table 5).

For both shoulder pain and shoulder disability, no significant differences ($p>0.05$) in the genotype/allele frequency distributions were noted between the no – low and moderate – high categories for the following SNPs: *ACAN* rs1126823 G>A, *ACAN* rs1516797 G>T, *ACAN* rs2882676 A>C, *BGN* rs1042103 G>A, *BGN* rs743641 A>T, *BGN* rs743642 G>T, *DCN* rs516115 C>T, *VCAN* rs2287926 G>A and *VCAN*

rs309559 A>G (Tables 4 and 5). The genotype distributions for the whole group were in Hardy–Weinberg equilibrium (HWE exact test $p>0.05$) for all SNPs investigated in this study (Supplementary Table 1).

Inferred haplotype frequency distributions between shoulder pain/disability categories

No significant differences were noted in the frequency distribution of the *ACAN* (rs1126823 G>A – rs1516797 G>T – rs2882676 A>C), *BGN* (rs1042103 G>A – rs743641 A>T – rs743642 G>T) or *VCAN* (rs11726 A>G – rs2287926 G>A – rs309559 A>G) inferred haplotypes between the no – low and moderate – high shoulder pain/disability categories ($p>0.05$) (Figure 1).

Inferred allele combination frequency distributions between shoulder pain/disability categories

No significant differences ($p>0.05$) were noted in the frequency distribution of the inferred *DCN* rs516115 C>T – *VCAN* (rs2287926 A>C – rs11726 A>G) allele combinations between no – low and moderate - high shoulder pain categories (Figure 2A). However, significant differences were noted in the frequencies of the *BGN* (rs743641 A>T – rs743642 G>T) – *VCAN* rs11726 A>G ($p=0.011$) inferred allele combinations between participants with no – low and moderate - high shoulder disability (Figure 2B). In particular, the T-T-G inferred allele combination of *BGN* (rs743641 A>T – rs743642 G>T) – *VCAN* rs11726 A>G was significantly over-represented ($p=0.002$; OR=2.347, 95% CI=1.215 – 4.534) in the moderate – high shoulder disability category compared to the no – low shoulder disability category (Figure 2B). Moreover, a trend was noted towards over-representation ($p=0.050$) of the T-T-A allele combination of *BGN* (rs743641 A>T – rs743642 G>T) – *VCAN* rs11726 A>G in the no – low shoulder disability category in comparison with the moderate to high disability category (Figure 2B).

Discussion

There is a paucity of studies investigating the role of genetic factors in modulating susceptibility to shoulder pain or disability amongst breast cancer survivors. Whereas proteoglycan gene polymorphisms or expression levels have been implicated in connective tissue conditions of the shoulder or other sites such as the knee [10, 12-16], their role in shoulder morbidity amongst breast cancer survivors is unknown. We found associations between *VCAN* 1429A>G genotype and *BGN* (rs743641 (251A>T) – rs743642 (318G>T)) – *VCAN* rs11726 (1429A>G) inferred allele combinations, and shoulder pain/disability among women following breast cancer treatment. To the best of our knowledge, our study is the first to report associations between proteoglycan gene polymorphisms and shoulder pain/disability amongst breast cancer survivors. This adds to the body of evidence indicating genetic predisposition to shoulder pain or disability following breast cancer treatment [7].

The *VCAN* rs11726 (1429A>G) polymorphism – associated with shoulder pain/disability both independently and in an inferred allele combination with *BGN*: rs743641 (251A>T) and rs743642 (318G>T) – is located in the 3' UTR gene region and has been shown to alter *VCAN* gene expression levels (Ensembl genome browser data, database version 100.38, Genome Reference Consortium Human Build 38) [33]. In particular, the G allele of *VCAN* rs11726 demonstrates higher levels of expression relative to the ancestral A allele in skeletal muscle and pancreatic cells [33]. The *VCAN* rs11726 polymorphism is also located within the long non-coding transcript: *VCAN-AS1*, which is an antisense RNA transcript for *VCAN* and hence another possible mechanism for regulating *VCAN* expression. Although non-significant, a trend towards under-representation of the G allele amongst participants reporting no – low levels of shoulder disability compared to those that reported moderate – high levels of shoulder disability was noted, indicating a protective effect ($p=0.069$, $OR=0.618$, $95\% CI=0.362 - 1.037$) (Table 5). The exact mechanism by which the *VCAN* rs11726 polymorphism may lead to shoulder disability or pain is not clear, given the versatile and complex functions of *VCAN* [17]. Possibly, *VCAN* interacts with important signaling factors such as the fibrogenesis factor TGF- β and the inflammatory factor nuclear factor kappa B (NF- κ B), which may contribute to shoulder morbidity by promoting fibrogenesis and nociceptive signaling, respectively [17, 34, 35]; Perhaps, increased expression of *VCAN* amongst G allele carriers leads to enhanced fibrosis and pain signaling in the shoulder in response to late treatment effects amongst breast cancer survivors. The *BGN* rs743641 and rs743642 polymorphisms are both located in the 3' UTR gene regulatory region, but have no reported gene expression correlations nor previous associations with connective tissue disorders of the shoulder. Another possible explanation for the role of the associated *VCAN* and *BGN* polymorphisms in this study could be that they are in linkage disequilibrium with other SNPs that are involved in the development of pain or disability. Whereas no previous associations have been reported between *VCAN* rs11726, *BGN* rs743641 or rs743642, and shoulder morbidity to date, other polymorphisms in *BGN* have been implicated in connective tissue disorders such as ACL ruptures [14, 16]. In addition, upregulation of *VCAN* and *BGN* expression has been demonstrated in a rat model of rotator cuff injury [36].

Consistent with previous reports on upper limb morbidity amongst breast cancer survivors [3, 6, 37], participants' age was significantly associated with both shoulder pain and shoulder disability following breast cancer treatment in our study. In logistic regression analyses using participants' age as a predictor, each 10 year increase in age was associated with a 10.1 % ($95\% CI=7.3 - 13.9\%$, $p=0.001$) and 6.5 % ($95\% CI= 4.0 - 9.0\%$, $p=0.011$) reduction in the odds of being in the moderate – high shoulder pain or disability category, respectively. The link between age and pain reporting remains unclear. One possible explanation is the reported reduction in pain sensitivity with age as determined from pressure pain threshold (PPT) measurements which may be relevant for movement-related pain that is measured by the SPADI instrument in our study [38]. Given that PPT measurements are subjective, the reported association may reflect changes in pain perception with age. Contrary to previous reports [2, 5, 6, 39, 40], type of breast surgery, having axillary surgery and receipt of adjuvant therapy were not significantly associated with shoulder pain or shoulder disability in our study. In fact, a higher frequency of the more aggressive surgical procedures mastectomy and ALND – compared to the conservative WLE and SLNB – was

observed in the no – low pain/disability group (Tables 2 and 3). This finding may perhaps be specific to our cohort; Pain reporting has been associated with ethnicity [1], and the average time after treatment in our cohort is longer than that of most similar studies. Consistent with our findings, De Groef *et al.* [41] reported a high prevalence of upper-limb morbidity amongst breast cancer patients who underwent the less invasive sentinel node-negative suggesting that upper limb morbidity among breast cancer survivors may not be largely explained by factors related to surgical management after long follow-up periods.

It was interesting to note the genotype associations of *ACAN*rs1126823 G>A, *ACAN*rs2882676 A>C, *BGN*rs743641 A>T, and *DCN*rs516115 C>T with treatment characteristics in our cohort. This may reflect, at least in part, the role of proteoglycans in the development and progression of cancer, thereby, influencing treatment options and selections. Although, no studies to date have demonstrated their roles in breast cancer, *BGN* and *DCN* have been implicated in the development and progression of endometrial, bladder, colon, blood or lung cancers [42].

Limitations

Despite being adequate for medium – large effect sizes ($OR \geq 2.0$), our sample size was largely underpowered (power <80%) for small effect sizes ($OR = 1.5$). As a result, we did not adjust for multiple comparisons based on the number of SNPs (familywise error rate), given its exploratory nature. Larger sample sizes may detect significant differences in other clinical/treatment characteristics and genotype/allele distributions included in this study. Although clinical relevance was used in creating shoulder pain/disability categories, there was no wide score gap between them. Therefore, close to the boundary score of 30, individuals with otherwise similar shoulder pain/disability characteristics may be in different categories. Ethnicity was determined by self-report, a less reliable method than genetic ancestry estimates and therefore, there is a possibility of undetermined population stratification in our sample. While determination of genetic ancestry is very useful in detecting population stratification, this study only tested targeted loci using functional polymorphisms in a hypothesis driven approach. Applying genetic ancestry estimates in this case would be a completely different study, much larger than the one described in this manuscript, such as a genome-wide association study. Lastly, associations between SNPs in *ACAN/BGN/DCN* genes and treatment characteristics (Supplementary Table 1), may have an undetermined influence on our findings and should be explored in greater depth with breast cancer risk.

Conclusion

Our findings provide evidence of association between polymorphisms in proteoglycan encoding genes and shoulder pain/disability among women following breast cancer treatment. Future studies in independent populations with larger sample sizes are warranted to replicate our findings and further characterize the reported associations.

Declarations

Acknowledgments

The authors would like to thank Ms Karryn Brown for assistance with genotyping. This study was funded by the National Research Foundation, South Africa.

Contributors

TSM contributed to genotyping, statistical analyses and manuscript preparation. AVS was responsible for genotyping QC, project development, analyses and manuscript preparation/editing. DS was responsible for project development, analyses and manuscript preparation/editing.

Funding

This study was funded in part by the University of Cape Town Research Council, the Medical Research Council and the National Research Foundation (NRF). TSM was funded by the University of Cape Town and the National Research Foundation.

Disclosure

Opinions expressed and conclusions arrived at, are those of the author and are not necessarily to be attributed to the funders. The authors declare no conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

1. Kramer, N., J. Ramjith, and D. Shamley, *Prevalence of shoulder morbidity after treatment for breast Cancer in South Africa*. Supportive Care in Cancer, 2019. **27**(7): p. 2591-2598.
2. Hidding, J.T., et al., *Treatment related impairments in arm and shoulder in patients with breast cancer: a systematic review*. PloS one, 2014. **9**(5): p. e96748-e96748.
3. Kootstra, J.J., et al., *A longitudinal study of shoulder and arm morbidity in breast cancer survivors 7 years after sentinel lymph node biopsy or axillary lymph node dissection*. Breast Cancer Research and Treatment, 2013. **139**(1): p. 125-134.
4. Zomkowski, K., et al., *Physical symptoms and working performance in female breast cancer survivors: a systematic review*. Disability and Rehabilitation, 2018. **40**(13): p. 1485-1493.
5. Shamley, D., et al., *Shoulder morbidity after treatment for breast cancer is bilateral and greater after mastectomy*. Acta Oncologica, 2012. **51**(8): p. 1045-1053.
6. Wang, L., et al., *Predictors of persistent pain after breast cancer surgery: a systematic review and meta-analysis of observational studies*. CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne, 2016. **188**(14): p. E352-E361.

7. Mafu, T.S., A.V. September, and D. Shamley, *KDR inferred haplotype is associated with upper limb dysfunction in breast cancer survivors of mixed ancestry*. Cancer management and research, 2019. **11**: p. 3829-3845.
8. Longo, U.G., et al., *Genetic basis of rotator cuff injury: a systematic review*. BMC medical genetics, 2019. **20**(1): p. 149-149.
9. Petrillo, S., et al., *Genetic factors in rotator cuff pathology: potential influence of col 5A1 polymorphism in outcomes of rotator cuff repair*. BMC medical genetics, 2020. **21**(1): p. 82-82.
10. Rahim, M., et al., *Genetics of musculoskeletal soft tissue injuries: Current status, challenges, and future directions*. 2019. p. 317-339.
11. Figueiredo, E.A., et al., *Rotator Cuff Tear Susceptibility Is Associated With Variants in Genes Involved in Tendon Extracellular Matrix Homeostasis*. Journal of Orthopaedic Research, 2020. **38**(1): p. 192-201.
12. Dean, B.J.F., S.L. Franklin, and A.J. Carr, *A systematic review of the histological and molecular changes in rotator cuff disease*. Bone & Joint Research, 2012. **1**(7): p. 158-166.
13. Chaudhury, S., et al., *Gene expression profiles of changes underlying different-sized human rotator cuff tendon tears*. Journal of Shoulder and Elbow Surgery, 2016. **25**(10): p. 1561-1570.
14. Ciężarczyk, P., et al., *Are genes encoding proteoglycans really associated with the risk of anterior cruciate ligament rupture?* Biology of sport, 2017. **34**(2): p. 97-103.
15. Willard, K., et al., *The interaction of polymorphisms in extracellular matrix genes and underlying miRNA motifs that modulate susceptibility to anterior cruciate ligament rupture*. Journal of Science and Medicine in Sport, 2018. **21**(1): p. 22-28.
16. Mannion, S., et al., *Genes encoding proteoglycans are associated with the risk of anterior cruciate ligament ruptures*. British Journal of Sports Medicine, 2014. **48**(22): p. 1640-1646.
17. Iozzo, R.V. and L. Schaefer, *Proteoglycan form and function: A comprehensive nomenclature of proteoglycans*. Matrix biology : journal of the International Society for Matrix Biology, 2015. **42**: p. 11-55.
18. Robinson, K.A., et al., *Decorin and biglycan are necessary for maintaining collagen fibril structure, fiber realignment, and mechanical properties of mature tendons*. Matrix biology : journal of the International Society for Matrix Biology, 2017. **64**: p. 81-93.
19. Wit, E., et al., *Genome-wide analysis of the South African Coloured population in the Western Cape*. Human genetics, 2010. **128**: p. 145-53.
20. Lahiri, D.K. and J.I. Nurnberger, Jr., *A rapid non-enzymatic method for the preparation of HMW DNA from blood for RFLP studies*. Nucleic acids research, 1991. **19**(19): p. 5444-5444.
21. MacDermid, J.C., P. Solomon, and K. Prkachin, *The Shoulder Pain and Disability Index demonstrates factor, construct and longitudinal validity*. BMC musculoskeletal disorders, 2006. **7**: p. 12-12.
22. Roach, K.E., et al., *Development of a shoulder pain and disability index*. Arthritis Care & Research, 1991. **4**(4): p. 143-149.

23. Gibbon, A., et al., *Defining the molecular signatures of Achilles tendinopathy and anterior cruciate ligament ruptures: A whole-exome sequencing approach*. PLOS ONE, 2018. **13**: p. e0205860.
24. Gauderman, W.J., *Sample size requirements for matched case-control studies of gene-environment interaction*. Stat Med, 2002. **21**(1): p. 35-50.
25. Dell-Inc., *Dell Statistica (Data Analysis Software System)*. 2016, Dell Inc.: Tulsa, OK 74104, USA.
26. RStudio-Team, *RStudio: Integrated Development Environment for R*. 2020, RStudio, PBC: Boston, MA.
27. R-Core-Team, *R: A language and environment for statistical computing*. 2020, R Foundation for Statistical Computing: Vienna, Austria.
28. Warnes, G., et al., *genetics: Population Genetics*
29. González, J.R., et al., *SNPassoc: an R package to perform whole genome association studies*. Bioinformatics, 2007. **23**(5): p. 654-655.
30. Sinnwell, J.P. and D.J. Schaid, *haplo.stats: Statistical Analysis of Haplotypes with Traits and Covariates when Linkage Phase is Ambiguous*. 2018.
31. Ripley, B.D. and W.N. Venables, *Modern Applied Statistics with S*. Fourth ed. 2002, New York: Springer.
32. Wickham, H., *ggplot2: Elegant Graphics for Data Analysis*. 2016, New York: Springer-Verlag.
33. Yates, A.D., et al., *Ensembl 2020*. Nucleic Acids Research, 2019. **48**(D1): p. D682-D688.
34. Wight, T.N., et al., *Versican—A Critical Extracellular Matrix Regulator of Immunity and Inflammation*. Frontiers in Immunology, 2020. **11**(512).
35. Sotoodehnejadmatalahi, F. and B. Burke, *Structure, function and regulation of versican: the most abundant type of proteoglycan in the extracellular matrix*. Acta Med Iran, 2013. **51**(11): p. 740-50.
36. Attia, M., et al., *Alterations of overused supraspinatus tendon: a possible role of glycosaminoglycans and HARP/pleiotrophin in early tendon pathology*. J Orthop Res, 2012. **30**(1): p. 61-71.
37. Hamood, R., et al., *Chronic pain and other symptoms among breast cancer survivors: prevalence, predictors, and effects on quality of life*. Breast Cancer Research and Treatment, 2018. **167**(1): p. 157-169.
38. El Tumi, H., et al., *Age-related changes in pain sensitivity in healthy humans: A systematic review with meta-analysis*. European Journal of Pain, 2017. **21**(6): p. 955-964.
39. Sagen, A., et al., *Upper Limb Physical Function and Adverse Effects After Breast Cancer Surgery: A Prospective 2.5-Year Follow-Up Study and Preoperative Measures*. Archives of Physical Medicine and Rehabilitation, 2014. **95**(5): p. 875-881.
40. Chrischilles, E.A., et al., *Upper extremity disability and quality of life after breast cancer treatment in the Greater Plains Collaborative clinical research network*. Breast Cancer Research and Treatment, 2019. **175**(3): p. 675-689.
41. De Groef, A., et al., *Arm lymphoedema and upper limb impairments in sentinel node-negative breast cancer patients: A one year follow-up study*. The Breast, 2016. **29**: p. 102-108.
42. Wei, J., et al., *Roles of Proteoglycans and Glycosaminoglycans in Cancer Development and Progression*. Int J Mol Sci, 2020. **21**(17).

Tables

Table 1: Inclusion and exclusion criteria for participant recruitment

Inclusion criteria	Exclusion criteria
>18 years old	History of shoulder or neck pathology prior to treatment for breast cancer
Females	Diagnosed connective tissue disorders such as rheumatoid arthritis or systemic lupus erythematosus
Unilateral breast cancer	Diagnosed renal insufficiency, diabetes mellitus or hypercholesterolemia
≥1 year after surgery	Diagnosed local recurrence
Self-declared "Mixed-ancestry" ethnicity	Diagnosed lymphedema

Table 2: Differences in demographic and clinical characteristics between shoulder pain categories

<i>Characteristic</i>	<i>Level/unit</i>	<i>No – Low Pain (n=183)</i>	<i>Moderate – High Pain (n=71)</i>	<i>p-value</i>	<i>Test</i>
<i>Age at consent</i>	years	60.8 (53.5 – 65.5)	53.8 (45.3 – 64.3)	0.001	MU
<i>Time after surgery</i>	years	3.0 (1.8 – 4.7)	2.3 (1.7 – 4.0)	0.139	MU
<i>Nodes removed</i>		10.0 (5.5 – 15.0)	9.0 (5.0 – 12.0)	0.154	MU
<i>Side of primary</i>	Left	52 (94)	54 (38)	0.779	F
	Right	48 (87)	46 (32)		
<i>Tumor grade</i>	III	24 (38)	23 (15)	0.930	χ^2
	II	51 (81)	53 (34)		
	I	26 (41)	23 (15)		
<i>Type of surgery</i>	Mastectomy	81 (148)	72 (50)	0.165	F
	WLE	19 (34)	28 (19)		
<i>Lymph node surgery</i>	ALND	85 (153)	78 (54)	0.263	F
	SLNB	15 (28)	22 (15)		
<i>Chemotherapy</i>	Yes	74 (131)	84 (58)	0.129	F
	No	26 (46)	16 (11)		
<i>Hormonal therapy</i>	Yes	77 (134)	79 (54)	0.863	F
	No	23 (39)	21 (14)		
<i>Hormonal regimen</i>	Tamoxifen	62 (106)	60 (41)	0.238	χ^2
	Aromatase inhibitor	8 (14)	4 (3)		
	Both	7 (12)	15 (10)		
<i>Radiotherapy</i>	Yes	68 (114)	74 (50)	0.531	F
	No	32 (53)	26 (18)		

Notes: Data presented as medians with interquartile ranges in parentheses, or % frequencies with actual counts (n) in parentheses. P-values in bold typeset indicate statistical significance ($p < 0.05$).

Abbreviations: WLE, Wide local excision; ALND, Axillary lymph node dissection; SLNB: Sentinel lymph node biopsy; MU: Mann Whitney U test; F: Fisher's exact test; χ^2 : Chi squared test.

Table 3: Differences in demographic and clinical characteristics between shoulder disability categories

<i>Characteristic</i>	<i>Level/unit</i>	<i>No – Low Disability (n=206)</i>	<i>Moderate – High Disability (n=48)</i>	<i>p-value</i>	<i>Test</i>
<i>Age at consent</i>	years	60.4 (53.2 – 65.2)	54.4 (45.0 – 64.9)	0.014	MU
<i>Time after surgery</i>	years	2.9 (1.7 – 4.5)	2.6 (1.8 – 4.3)	0.728	MU
<i>Nodes removed</i>		10.0 (5.0 – 14.0)	9.0 (6.0 – 13.0)	0.630	MU
<i>Side of primary</i>	Left	52 (106)	54 (26)	0.873	F
	Right	48 (97)	46 (22)		
<i>Tumor grade</i>	III	23 (41)	28 (12)	0.572	χ^2
	II	53 (96)	44 (19)		
	I	24 (44)	28 (12)		
<i>Type of surgery</i>	Mastectomy	79 (162)	77 (36)	0.693	F
	WLE	21 (42)	23 (11)		
<i>Lymph node surgery</i>	ALND	84 (170)	79 (37)	0.398	F
	SLNB	16 (33)	21 (10)		
<i>Chemotherapy</i>	Yes	76 (150)	81 (39)	0.567	F
	No	24 (48)	19 (9)		
<i>Hormonal therapy</i>	Yes	78 (151)	79 (37)	1.000	F
	No	22 (43)	21 (10)		
<i>Hormonal regimen</i>	Tamoxifen	62 (119)	60 (28)	0.452	χ^2
	Aromatase inhibitor	8 (15)	4 (2)		
	Both	8 (15)	15 (7)		
<i>Radiotherapy</i>	Yes	71 (134)	67 (30)	0.594	F
	No	29 (56)	33 (15)		

Notes: Data presented as medians with interquartile ranges in parentheses, or % frequencies with actual counts (n) in parentheses. P-values in bold typeset indicate statistical significance ($p < 0.05$).

Abbreviations: WLE, Wide local excision; ALND, Axillary lymph node dissection; SLNB: Sentinel lymph node biopsy; MU: Mann Whitney U test; F: Fisher's exact test; χ^2 : Chi squared test.

Table 4: Genotype and minor allele frequency distributions of the *ACAN*rs1126823 G>A, *ACAN*rs1516797 G>T, *ACAN*rs2882676 A>C, *BGN*rs1042103 G>A, *BGN*rs743641 A>T, *BGN*rs743642 G>T, *DCN*rs516115 C>T, *VCAN*rs11726 A>G, *VCAN*rs2287926 G>A and *VCAN*rs309559 A>G polymorphisms in participants with no–low pain and moderate–high pain in the shoulder following breast cancer treatment

Gene	SNP	Genotype or allele	No – low pain (n = 170)	Moderate – high pain (n = 61)	p-value
ACAN	rs1126823	G/G	29 (50)	33 (20)	0.609 (0.459)
		A/G	45 (76)	47 (29)	
		A/A	26 (44)	20 (12)	
		A	48 (164)	43 (53)	
	rs1516797	G/G	28 (47)	29 (18)	0.856 (0.959)
		G/T	44 (75)	46 (28)	
		T/T	28 (48)	25 (15)	
		T	50 (171)	48 (58)	
	rs2882676	A/A	32 (54)	27 (16)	0.724 (0.591)
		A/C	46 (78)	48 (29)	
		C/C	22 (37)	25 (15)	
		C	45 (152)	49 (59)	
BGN	rs1042103	G/G	74 (125)	70 (43)	0.350 (0.766)
		A/G	24 (41)	23 (14)	
		A/A	2 (4)	7 (4)	
		A	14 (49)	18 (22)	
	rs743641	A/A	48 (82)	46 (28)	0.801 (0.763)
		A/T	42 (71)	41 (25)	
		T/T	10 (17)	13 (8)	
		T	31 (105)	34 (41)	
	rs743642	G/G	54 (92)	52 (32)	0.476 (0.451)
		G/T	39 (67)	36 (22)	
		T/T	7 (11)	12 (7)	
		T	26 (89)	30 (36)	
DCN	rs516115	T/T	29 (49)	30 (18)	0.684 (0.572)
		C/T	53 (90)	48 (29)	
		C/C	18 (31)	23 (14)	

		C	45 (152)	47 (57)	0.751
VCAN	rs11726	G/G	29 (49)	48 (29)	0.004 (0.005)
		A/G	55 (94)	31 (19)	
		A/A	16 (27)	21 (13)	
		A	44 (148)	37 (45)	
	rs2287926	G/G	66 (112)	59 (36)	0.398 (0.374)
		A/G	29 (49)	38 (23)	
		A/A	5 (9)	3 (2)	
		A	20 (67)	22 (27)	
	rs309559	G/G	26 (44)	29 (18)	0.863 (0.877)
		A/G	49 (83)	46 (28)	
		A/A	25 (42)	25 (15)	
		A	49 (167)	48 (58)	

Notes: Genotype and allele frequencies are expressed as a percentage with the number of participants (n) in parentheses. P-values in bold typeset indicate significance ($p < 0.05$), whereas p-values in parentheses are adjusted for participants' age at consent.

Abbreviations: *ACAN*, Aggrecan; *BGN*, Biglycan; *DCN*, Decorin; *VCAN*, Versican.

Table 5: Genotype and minor allele frequency distributions of the *ACAN*rs1126823 G>A, *ACAN*rs1516797 G>T, *ACAN*rs2882676 A>C, *BGN*rs1042103 G>A, *BGN*rs743641 A>T, *BGN*rs743642 G>T, *DCN*rs516115 C>T, *VCAN*rs11726 A>G, *VCAN*rs2287926 G>A and *VCAN*rs309559 A>G polymorphisms in participants with no–low disability and moderate–high disability in the shoulder following breast cancer treatment

Gene	SNP	Genotype or allele	No – low disability (n = 188)	Moderate – high disability (n = 43)	p-value
ACAN	rs1126823	G/G	30 (56)	33 (14)	0.870 (0.987)
		A/G	46 (87)	42 (18)	
		A/A	24 (45)	26 (11)	
		A	47 (177)	47 (40)	
	rs1516797	G/G	29 (54)	26 (11)	0.864 (0.802)
		G/T	45 (84)	44 (19)	
		T/T	27 (50)	30 (13)	
		T	49 (184)	52 (45)	
	rs2882676	A/A	31 (58)	29 (12)	0.895 (0.861)
		A/C	46 (86)	50 (21)	
		C/C	23 (43)	21 (9)	
		C	46 (172)	46 (39)	
BGN	rs1042103	G/G	73 (137)	72 (31)	0.416 (0.401)
		A/G	25 (46)	21 (9)	
		A/A	3 (5)	7 (3)	
		A	15 (56)	17 (15)	
	rs743641	A/A	49 (92)	42 (18)	0.431 (0.442)
		A/T	42 (78)	42 (18)	
		T/T	10 (18)	16 (7)	
		T	30 (114)	37 (32)	
	rs743642	G/G	54 (102)	51 (22)	0.101 (0.101)
		G/T	40 (75)	33 (14)	
		T/T	6 (11)	16 (7)	
		T	26 (97)	33 (28)	
DCN	rs516115	T/T	29 (54)	30 (13)	0.976 (0.888)
		C/T	52 (97)	51 (22)	
		C/C	20 (37)	19 (8)	
		C	46 (171)	44 (38)	

VCAN	rs11726	G/G	30 (56)	51 (22)	0.024 (0.033)
		A/G	53 (99)	33 (14)	
		A/A	18 (33)	16 (7)	
		A	44 (165)	33 (28)	
	rs2287926	G/G	67 (125)	54 (23)	0.113 (0.118)
		A/G	28 (53)	44 (19)	
		A/A	5 (10)	2 (1)	
		A	19 (73)	24 (21)	
	rs309559	G/G	27 (51)	26 (11)	0.662 (0.685)
		A/G	49 (92)	44 (19)	
		A/A	24 (44)	30 (13)	
		A	48 (180)	52 (45)	

Notes: Genotype and allele frequencies are expressed as a percentage with the number of participants (n) in parentheses. P-values in bold typeset indicate significance ($p < 0.05$), whereas p-values in parentheses are adjusted for participants' age at consent.

Abbreviations: *ACAN*, Aggrecan; *BGN*, Biglycan; *DCN*, Decorin; *VCAN*, Versican.

Figures

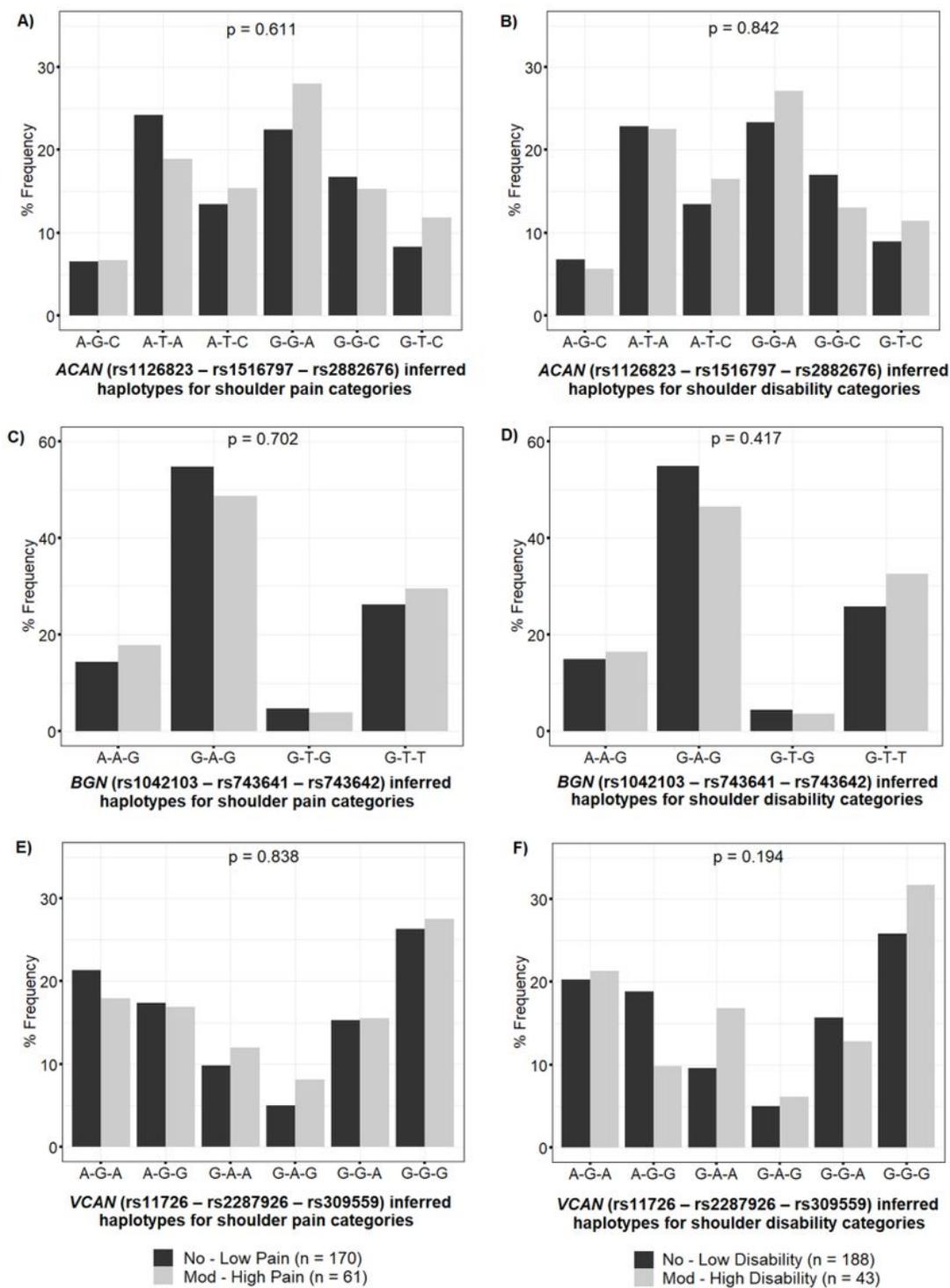


Figure 1

No significant differences were noted in the frequency distribution of the ACAN (rs1126823 G>A – rs1516797 G>T – rs2882676 A>C), BGN (rs1042103 G>A – rs743641 A>T – rs743642 G>T) or VCAN (rs11726 A>G – rs2287926 G>A – rs309559 A>G) inferred haplotypes between the no – low and moderate – high shoulder pain/disability categories ($p > 0.05$)

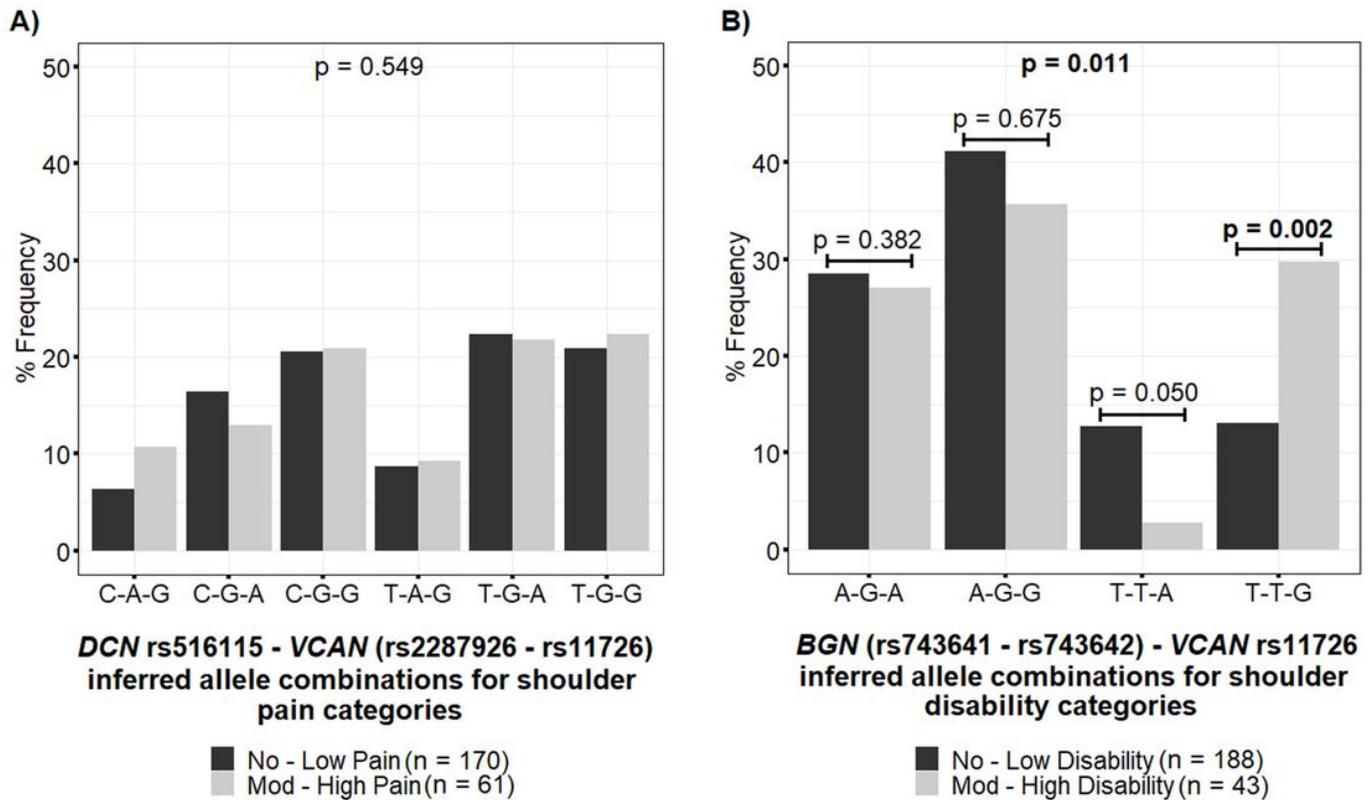


Figure 2

No significant differences ($p > 0.05$) were noted in the frequency distribution of the inferred DCN rs516115 C>T – VCAN (rs2287926 A>C – rs11726 A>G) allele combinations between no – low and moderate - high shoulder pain categories (Figure 2A). However, significant differences were noted in the frequencies of the BGN (rs743641 A>T – rs743642 G>T) – VCAN rs11726 A>G ($p = 0.011$) inferred allele combinations between participants with no – low and moderate - high shoulder disability (Figure 2B). In particular, the T-T-G inferred allele combination of BGN (rs743641 A>T – rs743642 G>T) – VCAN rs11726 A>G was significantly over-represented ($p = 0.002$; OR=2.347, 95% CI=1.215 – 4.534) in the moderate – high shoulder disability category compared to the no – low shoulder disability category (Figure 2B). Moreover, a trend was noted towards over-representation ($p = 0.050$) of the T-T-A allele combination of BGN (rs743641 A>T – rs743642 G>T) – VCAN rs11726 A>G in the no – low shoulder disability category in comparison with the moderate to high disability category

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplementarytable1.docx](#)