PLOS GENETICS

A genome-wide association study identifies 5 loci associated with frozen shoulder and implicates diabetes as a causal risk factor

Harry D. Green[1‡], Alistair Jones[2‡], Jonathan P. Evans[2], Andrew R. Wood[1], Robin N. Beaumont[1], Jessica Tyrrell[1], Timothy M. Frayling[1], Christopher Smith[2‡], Michael N. Weedon[2‡]*

1 Genetics of Complex Traits, University of Exeter Medical School, Exeter, United Kingdom, 2 Shoulder Unit, Princess Elizabeth Orthopaedic Centre, Royal Devon and Exeter Hospital, Exeter, United Kingdom

‡ HDG and AJ share equal first authorship on this work. CS and MNW share joint last authorship on this work.
* m.n.weedon@exeter.ac.uk

Abstract

Frozen shoulder is a painful condition that often requires surgery and affects up to 5% of individuals aged 40–60 years. Little is known about the causes of the condition, but diabetes is a strong risk factor. To begin to understand the biological mechanisms involved, we aimed to identify genetic variants associated with frozen shoulder and to use Mendelian randomization to test the causal role of diabetes. We performed a genome-wide association study (GWAS) of frozen shoulder in the UK Biobank using data from 10,104 cases identified from inpatient, surgical and primary care codes. We used data from FinnGen for replication and meta-analysis. We used one-sample and two-sample Mendelian randomization approaches to test for a causal association of diabetes with frozen shoulder. We identified five genome-wide significant loci. The most significant locus (lead SNP rs28971325; OR = 1.20, [95% CI: 1.16–1.24], p = 5x10^{-29}) contained *WNT7B*. This variant was also associated with Dupuytren's disease (OR = 2.31 [2.24, 2.39], p<1x10^{-300}) as were a further two of the frozen shoulder associated variants. The Mendelian randomization results provided evidence that type 1 diabetes is a causal risk factor for frozen shoulder (OR = 1.03 [1.02–1.05], p = 3x10^{-6}). There was no evidence that obesity was causally associated with frozen shoulder, suggesting that diabetes influences risk of the condition through glycemic rather than mechanical effects. We have identified genetic loci associated with frozen shoulder. There is a large overlap with Dupuytren's disease associated loci. Diabetes is a likely causal risk factor. Our results provide evidence of biological mechanisms involved in this common painful condition.

Author summary

Frozen shoulder is a painful condition that often requires surgery and affects up to 5% of individuals aged 40–60 years. Little is known about the causes but it is known to be more common in people with diabetes. In a dataset of 500,000 people, we used a genome-wide

association study to find genetic causes and a genetic technique called Mendelian randomisation to test if diabetes causes frozen shoulder. We found five new genetic variants that associate with frozen shoulder, and showed genetic overlap with Dupuytren's Disease, a similar condition that affects the fingers. We found an association with diabetes and obesity, but that the obesity association disappeared when we accounted for diabetes status, suggesting the condition is glycaemic rather than mechanical. Our Mendelian randomisation study showed evidence that type 1 diabetes has a causal effect on development of frozen shoulder, likely through a pathway involving long-term high blood glucose levels.

Introduction

Frozen shoulder, also known as adhesive capsulitis, affects 2–5% of the population at some point in their lives [1,2]. It is characterised by initial shoulder pain followed by a gradual reduction in range of movement or "freezing". Pain subsides but joint stiffness can continue for years, causing significant disability [3]. The onset of frozen shoulder is usually between 40 and 60 years [4].

Little is known about the causes of frozen shoulder. Genome-wide association studies can provide new insights into underlying biological mechanism and possible drug targets. For example, a recent GWAS of osteoarthritis has been used to identify new therapeutic targets for the condition [5]. To date, there have been no published genome wide association studies (GWAS) of frozen shoulder.

Diabetes is the strongest known risk factor for frozen shoulder. Individuals with diabetes have a greatly increased lifetime risk, with a hazard ratio of 1.33 [6]. Whether diabetes causes frozen shoulder is unclear because the association may reflect residual confounding by other risk factors such as age, obesity [7] and Dupuytren's disease [8].

Mendelian randomization is a statistical method that can be used to infer causal relationships between an exposure and an outcome by using genetic variants associated with the exposure [9]. The exposure associated variants can be used as an unconfounded proxy for the exposure, as their inheritance is random at conception. This method is now extensively used to infer causal associations and proof of principle examples include evidence that increased BMI causes diabetes [10] and increased LDL cholesterol causes coronary artery disease [11].

Using the UK Biobank, we performed a genome-wide association study (GWAS) to identify genetic variants associated with frozen shoulder. We used publicly available summary statistics from 176,899 individuals from the FinnGen study for replication [12]. We then used Mendelian randomization to test for causal associations of diabetes and obesity with frozen shoulder.

Methods

Ethics statement

Ethics approval for the UK Biobank study was obtained from the North West Centre for Research Ethics Committee (11/NW/0382)[13]. Written informed consent was obtained from all participants.

Frozen shoulder cases in the UK Biobank

The primary GWAS was performed on cases of frozen shoulder identified from the ICD-10 code M750, OPCS4 code W871 and read codes (N210, XE1FL and XE1Hm) from the primary

care data in UK Biobank. Controls were defined as individuals without a record of one of these codes.

Other phenotypes

ICD-10 codes were used for related disorders (Dupuytren's disease (M720), rotator cuff (M751), and calcific tendinitis of shoulder (M753)). Further, OPCS in Supplementary Table 1 of [14] (all sections except anatomical codes) and primary care read codes 7H320, 7H340, N236., Xa8RB, XE07l, XE1Fj, XM0ug, and XM1HO were used for Dupuytren's disease. Type 1 diabetes was classified favouring specificity over sensitivity, defined by diagnosis aged $\leq$ 20, on insulin within 1 year of diagnosis and at time of recruitment, not using oral antihyperglycemic agents, and not self-reporting type 2 diabetes, as described in [15]. Type 2 diabetes was defined by participants answering yes to the question 'Has a doctor ever told you that you have diabetes?' excluding those who reported using insulin within one year of diagnosis, were diagnosed under the age of 35, or were diagnosed within the past year [16]. We used UK Biobank variable 21001 for BMI measurement. HbA1c in mmol/mol is defined using variable n_30750_0_0 in UK Biobank. We defined diabetic retinopathy using self-report code 1276, diabetic neuropathy / ulcers using 1468 and diabetic nephropathy using 1607. Diabetes duration was defined by the difference between age at baseline and data-field 2976 (age diabetes diagnosed).

Genome-wide association study

We performed two case-control genome-wide association studies using BOLT-LMM [17], which applies a linear mixed model to test for an association between each SNP and the outcome trait, including population structure as a part of the model. Age, sex, study centre and genotyping chip were also included as covariates. We performed our primary GWAS using ICD10, OPCS and primary care codes, and a sensitivity analysis including only cases identified in ICD10 and OPCS. Briefly, principal component analysis was performed using individuals from the 1000 Genomes Project prior to projection of UK Biobank individuals into the principal component space. K-means clustering was subsequently applied to classify individuals as European, with centres initiated to the mean principal component values of each 1000 Genomes sub-population. The first 4 principal components were used in this analysis. We used LocusZoom to plot the resulting significant locus [18] and GTEx V8 to explore the possibility of functional variants [19]. P-values $< 5 \times 10^{-8}$ were considered significant for the GWAS. Betas and standard errors from BOLT-LMM were converted to log-odds ratios using $\log(OR) = \beta/(\mu(1-\mu))$, where μ is the case-control ratio, as per the BOLT-LMM user manual [17]. Standard errors were also divided by $\mu(1-\mu)$.

Replication and meta-analysis

We used summary statistics from freeze 4 of the FinnGen publicly available resource [12] to replicate the genome-wide significant findings in the discovery GWAS. Cases of frozen shoulder were defined in FinnGen using the ICD-10 code M750. METAL was used to perform a genome-wide meta-analysis of UKBB and FinnGen results, based on log-OR and standard errors [20]. We performed two meta-analyses using UK Biobank and FinnGen, the first using all cases from UK Biobank and the second using ICD-10 and OPCS codes.

Observational analyses

Observational analyses were performed on a subset of 379,708 unrelated individuals, 2,132 cases and 377,576 controls. We used a KING Kinship[21] to exclude those third-degree

relatives or closer. An optimal list of unrelated individuals was generated by preferentially removing individuals with the maximum number of relatives to allow maximum numbers of individuals to be included.

We tested for associations with demographic and clinical features using logistic regression models, firstly using a univariable logistic regression model on 7 variables of interest (sex, age, TDI, type 1 diabetes, type 2 diabetes, BMI and WHR) and secondly using all variables in a multivariable logistic regression model to identify independent associations. We analysed Type 1 and Type 2 diabetes separately. $p < 0.007$ (0.05/7) was considered significant. We performed a separate analysis specifically of diabetes-related variables (type, duration, retinopathy, neuropathy and nephropathy) for which $p < 0.01$ (0.05/5) was considered significant.

Mendelian randomization

We applied two different methods. The one sample Mendelian randomization (MR) results were performed in two stages: first, the association between each exposure and a genetic risk score was used to derive a genetically predicted exposure value, and second, these predicted values were used in a logistic regression model of frozen shoulder, adjusting for ancestral principal components, assessment centre, genotyping platform, age and sex.

Second, we used two-sample methods, which involve regressing the effect sizes of variant–outcome associations against the effect sizes of the variant–risk factor associations for a set of SNPs associated with the exposure. We performed inverse variance weighted (IVW) instrumental variable analysis and two further methods that are more robust to potential violations of the standard MR assumptions (MR-Egger [22] and weighted-median MR [23]). We used effect sizes and 30 SNPs for type 1 diabetes reported in [24] using the Wellcome Trust Case Control Consortium, and a type 2 diabetes genetic risk score of 88 variants identified in [25,26]. The T1D-GRS uses 2 SNPs, rs2187668 and rs7454108, to tag and categorize the high risk DR3/DR4 haplotypes [24], this requires access to individual level data which we did not have for the FinnGen study. We therefore performed two analyses: one where we used the 2 SNPs to impute DR3/DR4 status and another where we used the SNPs as standard risk SNPs. We performed the above two-sample Mendelian randomisation in UK Biobank cases identified by ICD10 and OPCS codes, all UK Biobank cases and the FinnGen cohort. We also performed Mendelian-Randomisation using log-odds ratios and standard errors from the two genome-wide meta-analyses.

P-values < 0.01 were considered significant for Mendelian randomisation results. Mendelian randomisation results using a diabetes status as an exposure should be interpreted in terms of liability, which may cause frozen shoulder via changes in diabetes status and / or by independent pathways such as glycaemia[27].

Results

Diabetes is associated with frozen shoulder in the UK Biobank

To identify associated risk factors for frozen shoulder we tested several measures of diabetes and its related traits to assess their independent contributions. There were 2,540 cases of frozen shoulder in the UK Biobank based on ICD-10 and OPCS codes, and 10,104 including GP records. A Venn diagram showing overlap can be found in **S1 Fig**. Demographic and clinical associations with frozen shoulder defined by ICD10 and OPCS can be found in **Table 1**. Those with frozen shoulder were more likely female, had lower Townsend deprivation index, more likely obese (higher BMI and waist-hip-ratio) and more likely to have either type 1 or type 2 diabetes, with type 1 giving the strongest association. In a multivariable logistic regression model, only sex, types 1 and type 2 diabetes showed association.

Table 1. Demographic and clinical associations with frozen shoulder in the UK Biobank. The OR and p value columns were calculated using univariable logistic regression. The adjusted OR and adjusted p value columns were calculated using a multivariable logistic regression model. Adjusted results for diabetes types were calculated from a model excluding the other type. These analyses used only unrelated individuals in the UK Biobank.

Variable	Cases (2,132)	Controls (377,576)	OR [95% CI]	p value	Adjusted OR [95% CI]	Adjusted p value
Male sex	174,125 (46%)	820 (38%)	0.73 [0.67–0.80]	2×10^{-12}	0.64 [0.58–0.71]	1×10^{-18}
Age [SD]	57.2 [8.0]	57.2 [7.4]	1.00 [0.99–1.01]	0.86	1 [1.00–1.00]	0.59
Townsend deprivation index [SD]	-1.26 [3.1]	-1.48 [3.0]	1.02 [1.01–1.04]	7×10^{-4}	1.01 [0.99–1.03]	0.10
Type 1 diabetes	23 (1.2%)	356 (0.09%)	12.23 [8.00–18.69]	9×10^{-31}	13.09 [8.55–20.03]	2×10^{-32}
Type 2 diabetes	179 (9.0%)	11,887 (3.2%)	2.99 [2.56–3.49]	1×10^{-43}	3.00 [2.54–3.54]	6×10^{-39}
BMI [SD]	28.0 [4.9]	27.4 [4.8]	1.12 [1.07–1.18]	1×10^{-10}	1.03 [0.97–1.09]	0.12
WHR [SD]	0.873 [0.09]	0.871 [0.09]	1.16 [1.11–1.21]	1×10^{-17}	1.03 [0.97–1.09]	0.09

https://doi.org/10.1371/journal.pgen.1009577.t001

Diabetes duration was nominally associated with frozen shoulder adjusted for diabetes type (OR per additional year with diabetes 1.02 [1.00–1.04], p = 0.02). HbA1c associated with frozen shoulder independently of type or duration of diabetes (OR = 1.09 [95% CI: 1.02–1.17], p = 0.009). Diabetic retinopathy (OR = 1.99 [95% CI, 1.19–3.33], p = 0.009) associated with frozen shoulder independently of type or duration of diabetes. This suggests that individuals with longer duration and less well controlled diabetes have higher risk of frozen shoulder.

Three loci are associated with frozen shoulder from discovery GWAS

Fig 1 presents the results of the frozen shoulder genome-wide association study using all frozen shoulder cases. We identify a genome-wide significant peak on chromosome 22. The A allele at the lead SNP, rs28971325, has a frequency of 26.4% in cases and 23.2% in controls (OR = 1.20, [95% CI: 1.16–1.24], p = 5×10^{-29}). We also observe two additional genome-wide
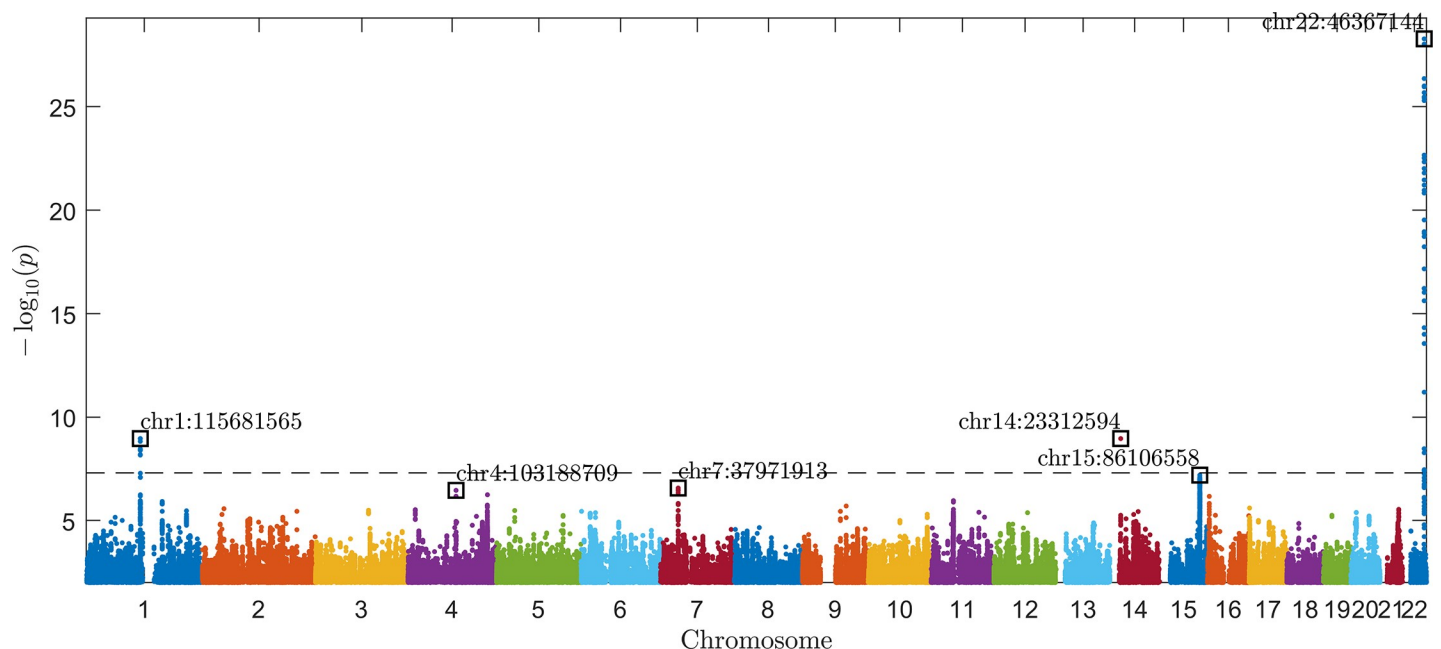
Fig 1. Manhattan plot of discovery GWAS for frozen shoulder in UK Biobank. The plot shows $-\log 10(p)$ values for each single nucleotide polymorphism [SNP] in the HRC Imputation Panel and their association with frozen shoulder defined by ICD10, OPCS codes and GP records with p < 0.01, computed using BOLT-LMM and plotted using an in-house MATLAB script which we have made publicly available on the MATLAB File Exchange [31]. The horizontal dashed line is the genome-wide significance threshold at p = 5×10^{-8}. Positions are based on the hg19 reference human genome.

https://doi.org/10.1371/journal.pgen.1009577.g001

significant signals: rs5777216 on chromosome 1 (OR = 0.92 [95%CI 0.89–0.94], p = 1x10^{-9}) and rs1042704 on chromosome 14 (OR = 1.11 [95%CI 1.07–1.15], p = 1x10^{-9}).

When including only ICD10 and OPCS codes in the analysis, rs28971325 associated with OR = 1.32 (95%CI 1.22–1.41), p = 7x10^{-17}. rs5777216 (OR = 0.91 [95%CI 0.86–0.96], p = 1x10^{-3}) and rs1042704 (OR = 1.12 [95%CI 1.05–1.20], p = 8x10^{-4}) were not genome-wide significant. A Manhattan Plot showing these results can be found in S2 Fig.

Replication and meta-analyses of genome-wide association studies

We sought replication of the association signals in the FinnGen study. FinnGen have recently released GWAS summary statistics for 1801 diseases and traits, including frozen shoulder [12]. The lead SNP from the UK Biobank, rs28971325, was associated with frozen shoulder in Finn-Gen with a similar effect size to the UK Biobank results using ICD10 and OPCS codes (OR = 1.30 [95%CI 1.20–1.41], p = 6x10^{-10}). The frequency of the A allele was 20.8% in cases compared to 17.6% in controls. rs1042704 was nominally associated with frozen shoulder in the FinnGen cohort (OR = 1.15 [95%CI: 1.07–1.24] p = 2x10^{-4}), while rs5777216 (OR = 0.94 [95%CI: 0.89–1.00] p = 6x10^{-2}) was not.

We note five genome-wide significant loci when performing a meta-analysis using all UK Biobank cases and FinnGen, including the three previously associated loci. Odds ratios for these SNPs were similar across all three definitions, except for rs28971325, which was weaker in UKBB's GP records. The summary statistics can be found in Table 2. Manhattan Plots of the meta-analyses can be found in S3 and S4 Figs.

WNT7B, MMP14 and SFRP4 are potential causal genes

S5–S9 Figs provide LocusZoom plots [18] for the 5 associated loci. For the strongest associated locus, rs28971325 (S5 Fig), there were no coding variants with an r^2 > 0.8. There are no gene expression associations in GTEX for rs28971325. However, a recent study found *WNT7B* was the second most differentially expressed transcript genome-wide with a log fold change of 7 (*p* = 1x10^{-16}) in anterior capsule tissue from 22 patients undergoing arthroscopic capsulotomy surgery for frozen shoulder compared to 26 controls [28]. The lead SNP at the next strongest associated locus, rs1042704, is a missense variant in the *MMP14* gene (S6 Fig). This variant has previously been associated with Dupuytren's disease and has been shown to have a 99% posterior probability of being the causal variant [29]. The chromosome 7 locus is also a known Dupuytren's locus, and previous work has suggested *SFRP4* is the causal gene [29].

Frozen shoulder associated SNPs have been associated with Dupuytren's contracture and bone mineral density

A previous GWAS study of Dupuytren's disease found an association in the same genomic region of *WNT7B* [29]. The lead SNP for Dupuytren's disease, rs7291412, was not associated

Table 2. Association of lead SNPs in the full meta analysis in the two input GWAS results (ICD+OPCS+GP in UKBB and FinnGen) and the UKBB result excluding GP records. The effect allele frequency is from the UK Biobank cohort.

SNP	Effect Allele (Freq)	Meta Analysis	ICD+OPCS	ICD+OPCS+GP	FinnGen
22:46367144 rs28971325	A (0.233)	1.21 (1.25–1.18) 8E-37	1.32 (1.24–1.41) p = 7E-17	1.20 (1.16–1.24) p = 5E-29	1.30 (1.20–1.41) p = 6E-10
14:23312594 rs1042704	A (0.215)	1.12 (1.15–1.08) 1E-12	1.12 (1.05–1.20) p = 8E-04	1.11 (1.07–1.15) p = 1E-09	1.15 (1.07–1.24) p = 2E-04
1:115681565 rs5777216	GT (0.403)	1.09 (1.11–1.06) 3E-10	0.91 (0.86–0.96) p = 1E-03	0.92 (0.89–0.94) p = 1E-09	0.94 (0.89–1.00) p = 6E-02
15:86125607 rs17570529	T (0.266)	0.92 (0.94–0.89) 3E-09	1.10 (1.03–1.17) p = 4E-03	1.09 (1.05–1.12) p = 1E-07	1.10 (1.03–1.18) p = 6E-03
7:37981961 rs2472660	C (0.355)	0.93 (0.95–0.90) 2E-08	1.06 (1.00–1.12) p = 4E-02	1.07 (1.04–1.10) p = 6E-06	1.13 (1.06–1.20) p = 3E-04

Table 3. Association of frozen shoulder genetic loci with Dupuytren's Disease.

CHR	SNP	BP	OR	95%CI	P
22	rs28971325	46367144	2.31	2.24–2.39	$<1 \times 10^{-300}$
14	rs1042704	23312594	1.18	1.14–1.22	3×10^{-21}
1	rs5777216	115681565	1.02	0.99–1.05	0.21
15	rs17570529	86125607	1.08	1.04–1.11	1×10^{-5}
7	rs2472660	37981961	1.32	1.28–1.36	7×10^{-73}

https://doi.org/10.1371/journal.pgen.1009577.t003

with frozen shoulder in UK Biobank (OR = 0.98 [95%CI: 0.95–1.01], p = 0.17) or FinnGen (OR = 1.00 [95%: 0.94, 1.06], p = 0.92). Of the five genome-wide significant loci in the meta-analysis, rs28971325, rs1042704 and rs2472660 associated with Dupuytren's disease at the genome-wide significant level, rs17570529 was nominally significant (p = 1×10^{-5}), and rs2472660 did not associate (p = 0.21) (**Table 3**). None of these five loci associated with either rotator cuff or calcific tendinitis of shoulder at p<0.01.

Dupuytren's disease and other fibroblastic diseases

There was limited overlap between Frozen Shoulder and Dupuytren's disease observationally: 61 (2.4%) of frozen shoulder cases in the ICD10+OPCS definition and 198 (2.0%) including GP record cases had an ICD10 code for Dupuytren's disease. However, many of the previously reported Dupuytren's disease SNPs were associated with frozen shoulder with a correlation in effect sizes (R^2 = 0.72, p = 1×10^{-7}, **S10 Fig**). The association of the 5 SNPs with frozen shoulder did not change when we excluded individuals with Dupuytren's disease from the cases.

Mendelian randomization provides evidence that diabetes is causal to frozen shoulder

We used Mendelian randomization to explore whether the associations with Type 1 and Type 2 diabetes reported in **Table 1** were causal. Using 1-sample Mendelian randomization methods, genetic data provided evidence that type 1 diabetes causes frozen shoulder: OR 1.05 [95% CI: 1.02–1.09], p = 0.002 using ICD-10 and OPCS codes and OR 1.04 [95% CI: 1.02–1.06], p = 2×10^{-6} including primary care records. Evidence of a causal role of type 2 diabetes was weaker: OR 1.10 [95% CI: 0.99–1.22], p = 0.07 using ICD-10 codes and OR 1.07 [95% CI: 1.02–1.13], p = 0.006 including cases identified by UK Biobank primary care records. These results were consistent when we performed a sensitivity analysis adjusting both stages of the regression for significant variables in **Table 1** (BMI, WHR and TDI) to correct for potential pleiotropy (**S1 Table**).

The association with type 1 diabetes was replicated using more robust, two-sample MR methods IVW (p = 3×10^{-6}) and MR-Egger (p = 2×10^{-6}). There was limited evidence from the two-sample methods that type 2 diabetes causes frozen shoulder (IVW: p = 0.06, MR-Egger: p = 0.96). The supplementary tables contain the full results from our Mendelian randomization analyses, including T1D-GRS results that exclude HLA variants which have greater potential for pleiotropy, for which the effect sizes were consistent (**S1 Table** shows the two stage regression, **S2 Table** shows the IVW results and heterogeneity p values, and **S3 Table** shows the MR-Egger results). **Figs 2 and 3** show plots showing the lines of best fit for MR-Egger, IVW, Median IV and Penalised Median IV methods for type 1 and type 2 diabetes defined by ICD-10 codes, demonstrating in **Fig 2**: for type 1 diabetes SNPs, a strong correlation between the betas for type 1 diabetes and frozen shoulder and in **Fig 3**: no association for type 2 diabetes SNPs.
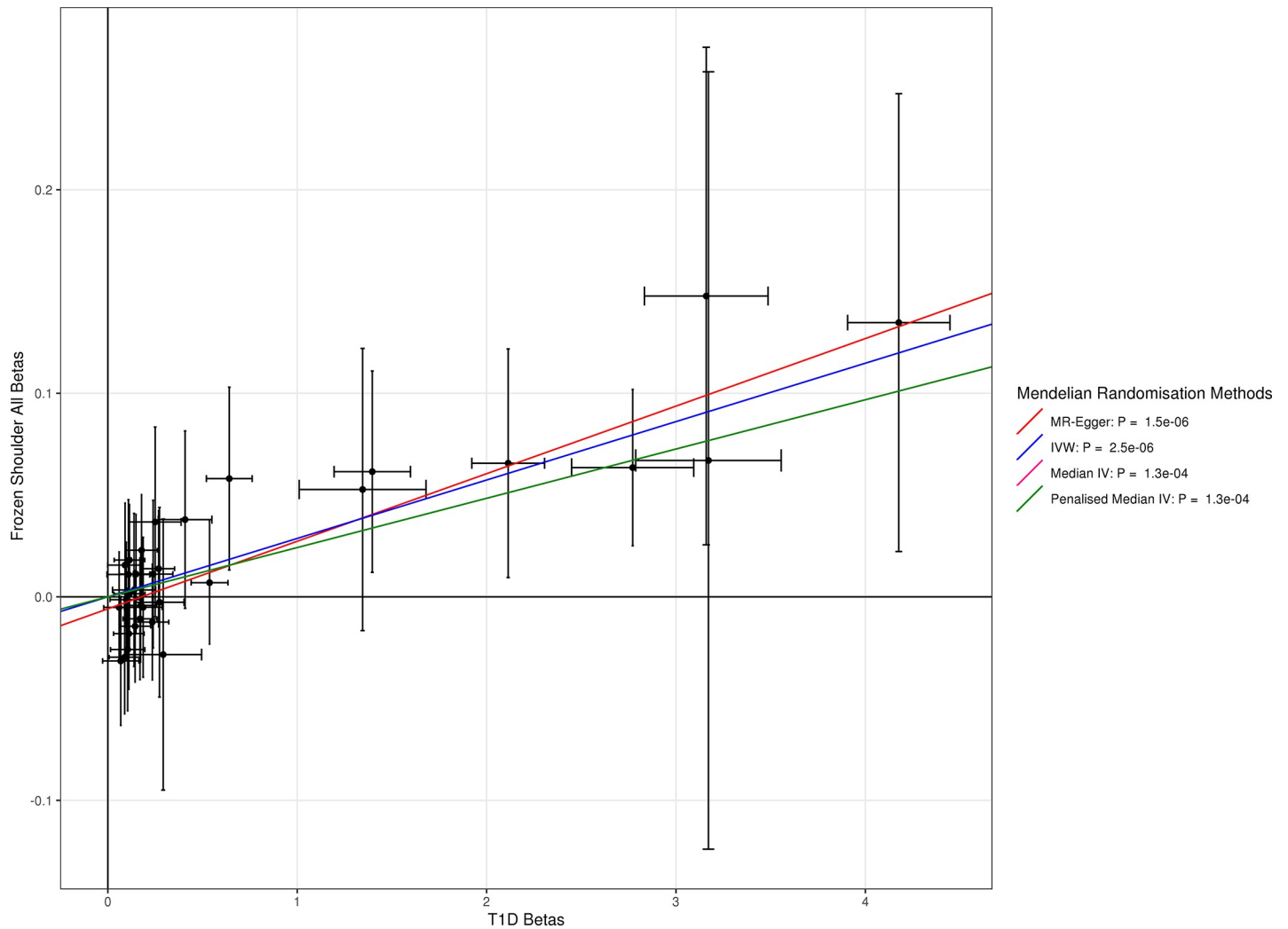
Fig 2. Two Sample Mendelian Randomization Results, showing for each SNP in the Type 1 Diabetes GRS in [20], the log-odds ratio for type 1 diabetes on the x axis and the log-odds ratio for frozen shoulder (defined by ICD-10, OPCS and GP records) on the y axis.

Discussion

We have identified robust associations of common genetic variants with frozen shoulder. Frozen shoulder is a condition which affects up to 5% of the population around between the ages of 40 to 60, but the causes of the disease, and particularly the transient nature of the condition are unknown. The genome-wide association study and Mendelian randomization analyses we report here provide new insights into the underlying causes of the condition.

WNT7B is a candidate causal gene at the most strongly associated locus. As with most GWAS studies, further work is needed to identify the causal variant at the locus. GTEX analyses did not identify any strong eQTL's with the lead SNP at the associated locus. However, a recent study performed RNA-seq on anterior capsule tissue from 22 patients undergoing arthroscopic capsulotomy surgery for adhesive capsulitis and compared to 26 undergoing arthroscopic stabilization surgery for a different condition [28]. *WNT7B* was the second most differentially expressed transcript genome-wide with a log fold change of 7 ($p = 1 \times 10^{-16}$), although it was noted that the expression levels of *WNT7B* was relatively low. The WNT
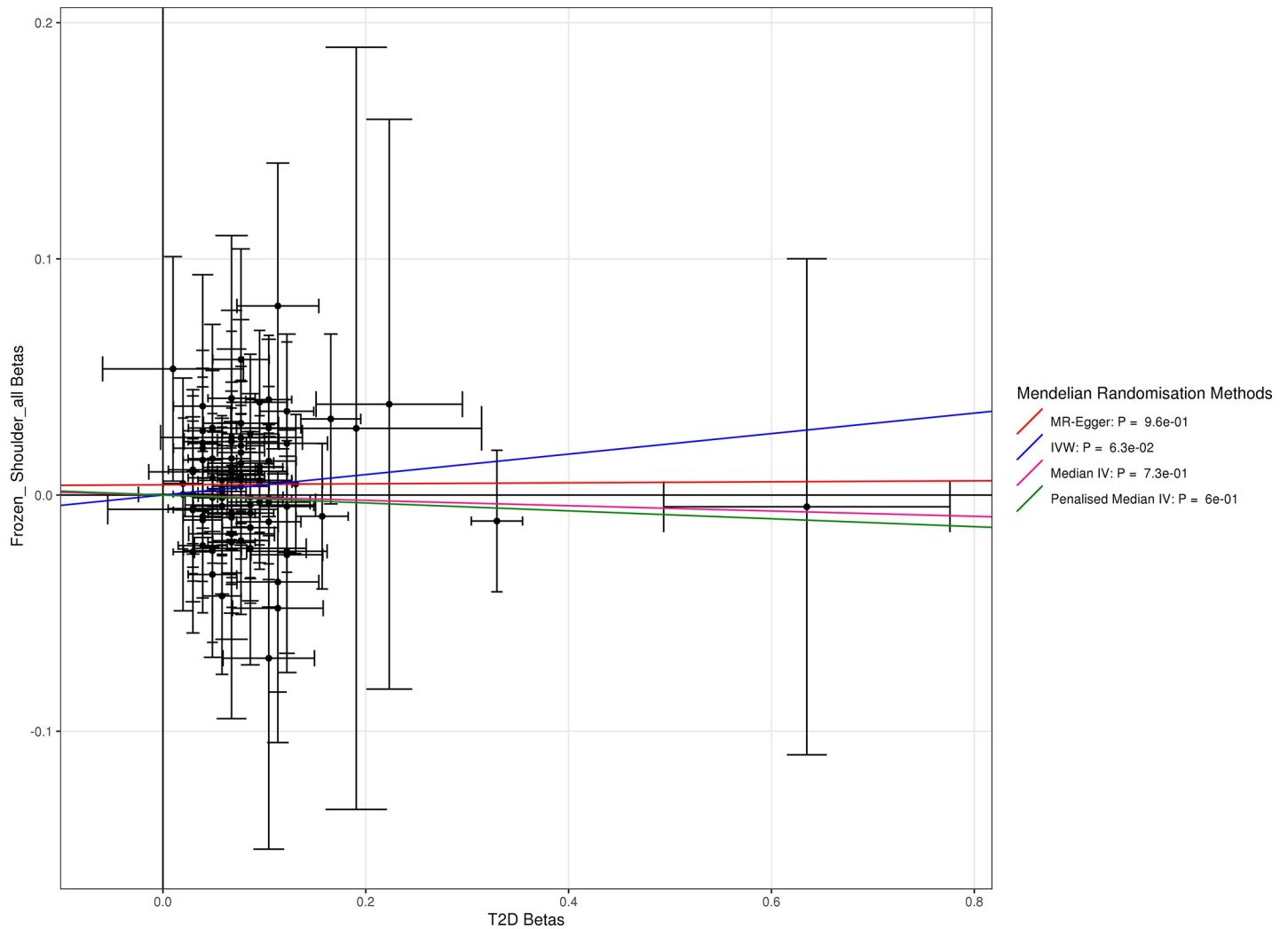
Fig 3. Two Sample Mendelian Randomization Results, showing for each SNP in the Type 1 Diabetes GRS in [20], the log-odds ratio for type 2 diabetes on the x axis and the log-odds ratio for frozen shoulder (defined by ICD-10, OPCS and GP records) on the y axis.

signalling pathway has been highlighted by previous GWAS of related fibroblastic diseases [30]. Other potential causal genes at associated loci included *MMP14*, a missense variant previously finemapped for Dupuytren's disease, and *SFRP4*.

Three of the identified loci for frozen shoulder also associate with Dupuytren's disease. Dupuytren's disease is a common condition which is characterised by a hand deformity where there is contracture of connective tissue within the palm and digits preventing full finger extension. A previous GWAS of Dupuytren's identified 26 associated loci with the condition [29]. One of the more strongly associated loci also contained the *WNT7B* locus. However, the lead SNP at that locus has not been associated with frozen shoulder and does not associate with frozen shoulder in the UK Biobank or FinnGen. The odds ratio for the lead SNP in our study has a significantly stronger odds ratio for Dupuytren's than the strongest association signal from the Dupuytren's GWAS. The explanation may be that older versions of SNP chips or imputation panels were used in the previous study which did not capture the lead SNP from our current study of frozen shoulder. Overall there is a strong overlap in signals between Dupuytren's disease and frozen shoulder. Some overlap might be expected, as both conditions

are a result of contracture of connective tissue planes. WNT is a known regulator of planar cell polarity and causes tissue planes to contract via the non-canonical pathway. Dupuytren's disease differs from frozen Shoulder in that frozen shoulder is most commonly a transient condition.

Diabetes is known to be an observational risk factor for frozen shoulder, but the causal nature of the association was unclear. Our Mendelian randomization analyses in UK Biobank provide evidence that Type 1 diabetes is causal for the condition. The weaker association with Type 2 diabetes is likely due to differences in duration of diabetes or earlier diabetes onset, as Type 1 diabetes is generally diagnosed earlier (half before the age of 30 years) whereas the diagnosis of Type 2 diabetes generally occurs later in life (after 50 years). It is likely all individuals with diabetes would have an increased risk of frozen shoulder, with those with longer duration (like those individuals with Type 1 diabetes) and worse glycaemic control at increased risk.

There are limitations to our study. The UK Biobank ICD-10 data is reliant on patients having a coded diagnosis at hospital, which is reliant on accuracy of hospital coding, and also reliant on the condition being severe enough for the patient to go to hospital. This may result in less serious cases being classified as controls. As such, they may be overly specific, while the primary care records are potentially overly sensitive and include cases that are not true frozen shoulder cases because it can be difficult to accurately diagnose in primary care. Consistent with this the effect size for frozen shoulder was smaller when cases were based on presence in primary care records rather than inpatient ICD-10 codes for the most associated loci. This study was performed using only white Europeans, and further study is needed to determine if results replicate for other ethnic groups. The UK Biobank only includes patients between the ages of 40 and 69 at recruitment, although this covers the most common incidence age range for frozen shoulder.

We have identified genetic loci associated with frozen shoulder. There is a large overlap with Dupuytren's disease genetic loci. Diabetes is a likely causal risk factor. Our results provide evidence of biological mechanisms involved in this common painful condition.

Supporting information

S1 Table. One sample Mendelian randomisation results. A table showing results from a two-stage regression where the GRS for the exposure was regressed against the exposure, and the genetically predicted exposure was used to predict frozen shoulder. Adjusted refers to a sensitivity analysis in which both stages of the regression were adjusted for the significant variables in Table 1.
(DOCX)

S2 Table. IVW Mendelian randomisation results. A table showing results the IVW analysis. Meta Analysis 1 refers to using the betas and standard errors from the meta-analysis GWAS with FinnGen using ICD10 + OPCS from UKBB. Meta Analysis 2 refers to the same using ICD10 + OPCS + GP records from UKBB. The P het column contains heterogeneity.
(DOCX)

S3 Table. MR-Egger results. A table showing results the MR-Egger results. Meta Analysis 1 refers to using the betas and standard errors from the meta-analysis GWAS with FinnGen using ICD10 + OPCS from UKBB. Meta Analysis 2 refers to the same using ICD10 + OPCS + GP records from UKBB.
(DOCX)

S1 Fig. Venn diagram of UK Biobank cases. This Venn diagram shows overlap between the different case definitions of frozen shoulder (ICD10, OPCS, and GP record codes) in the UK

Biobank.
(TIF)

S2 Fig. Manhattan plot of sensitivity analysis GWAS for frozen shoulder in UK Biobank.
The plot shows–log10(p) values for the association of each single nucleotide polymorphism
[SNP] in the HRC Imputation Panel and their association with UK Biobank frozen shoulder
cases defined by ICD10 and OPCS codes. The horizontal dashed line is the genome-wide significance threshold at $p = 5 \times 10^{-8}$.
(TIF)

S3 Fig. Manhattan plot of primary meta-analysis. The plot shows–log10(p) values for the
association of each single nucleotide polymorphism [SNP] in the HRC Imputation Panel and
their association in the meta-analysis using UK Biobank ICD10 + OPCS and FinnGen. The
horizontal dashed line is the genome-wide significance threshold at $p = 5 \times 10^{-8}$.
(TIF)

S4 Fig. Manhattan plot of secondary meta-analysis. The plot shows–log10(p) values for the
association of each single nucleotide polymorphism [SNP] in the HRC Imputation Panel and
their association in the meta-analysis using UK Biobank ICD10 + OPCS + GP Records and
FinnGen. The horizontal dashed line is the genome-wide significance threshold at $p = 5 \times 10^{-8}$.
(TIF)

S5 Fig. LocusZoom plot of rs28971325.
(TIF)

S6 Fig. LocusZoom plot of rs1042704.
(TIF)

S7 Fig. LocusZoom plot of rs5777216.
(TIF)

S8 Fig. LocusZoom plot of rs17570529.
(TIF)

S9 Fig. LocusZoom plot of rs2472660.
(TIF)

**S10 Fig. Scatterplot of Frozen Shoulder betas against Dupuytren betas for Dupuytrens
SNPs.** Betas and 95% CIs for association between Frozen Shoulder and between Dupuytren's
Disease for all Dupuytren's SNPs in Ng. et. al.
(TIF)

Author Contributions

Conceptualization: Alistair Jones, Jonathan P. Evans, Christopher Smith, Michael N.
    Weedon.

Formal analysis: Harry D. Green, Michael N. Weedon.

Funding acquisition: Timothy M. Frayling, Michael N. Weedon.

Investigation: Harry D. Green, Michael N. Weedon.

Methodology: Harry D. Green, Andrew R. Wood, Robin N. Beaumont, Jessica Tyrrell, Timothy M. Frayling, Michael N. Weedon.

Software: Harry D. Green, Andrew R. Wood, Robin N. Beaumont, Jessica Tyrrell, Michael N. Weedon.

Supervision: Timothy M. Frayling, Michael N. Weedon.

Validation: Harry D. Green.

Visualization: Harry D. Green.

Writing – original draft: Harry D. Green, Alistair Jones, Jonathan P. Evans, Andrew R. Wood, Robin N. Beaumont, Jessica Tyrrell, Timothy M. Frayling, Christopher Smith, Michael N. Weedon.

Writing – review & editing: Harry D. Green, Alistair Jones, Jonathan P. Evans, Andrew R. Wood, Robin N. Beaumont, Jessica Tyrrell, Timothy M. Frayling, Christopher Smith, Michael N. Weedon.

References

1. Morén-Hybbinette I, Moritz U, Scherstén B. The clinical picture of the painful diabetic shoulder—natural history, social consequences and analysis of concomitant hand syndrome. Acta Med Scand. 1987; 221:73–82. https://doi.org/10.1111/j.0954-6820.1987.tb01247.x PMID: 2436441

2. Zreik NH, Malik RA, Charalambous CP. Adhesive capsulitis of the shoulder and diabetes: a meta-analysis of prevalence. Muscles Ligaments Tendons J. 2016; 6:26–34. https://doi.org/10.11138/mltj/2016.6.1.026 PMID: 27331029

3. Hand C, Clipsham K, Rees JL, Carr AJ. Long-term outcome of frozen shoulder. J shoulder Elb Surg. 2008; 17:231–236. https://doi.org/10.1016/j.jse.2007.05.009 PMID: 17993282

4. Uppal HS, Evans JP, Smith C. Frozen shoulder: A systematic review of therapeutic options. World J Orthop. 2015; 6:263–268. https://doi.org/10.5312/wjo.v6.i2.263 PMID: 25793166

5. Tachmazidou I, Hatzikotoulas K, Southam L, Esparza-Gordillo J, Haberland V, Zheng J, et al. Identification of new therapeutic targets for osteoarthritis through genome-wide analyses of UK Biobank data. Nat Genet. 2019; 51:230–236. https://doi.org/10.1038/s41588-018-0327-1 PMID: 30664745

6. Huang Y-P, Fann C-Y, Chiu Y-H, Yen M-F, Chen L-S, Chen H-H, et al. Association of diabetes mellitus with the risk of developing adhesive capsulitis of the shoulder: a longitudinal population-based followup study. Arthritis Care Res (Hoboken). 2013; 65:1197–1202. https://doi.org/10.1002/acr.21938 PMID: 23281342

7. Kingston K, Curry EJ, Galvin JW, Li X. Shoulder adhesive capsulitis: epidemiology and predictors of surgery. J shoulder Elb Surg. 2018; 27:1437–1443. https://doi.org/10.1016/j.jse.2018.04.004 PMID: 29807717

8. Smith SP, Devaraj VS, Bunker TD. The association between frozen shoulder and Dupuytren's disease. J shoulder Elb Surg. 2001; 10:149–151. https://doi.org/10.1067/mse.2001.112883 PMID: 11307078

9. Davies NM, Holmes M V, Davey Smith G. Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians. BMJ. 2018; 362:k601. https://doi.org/10.1136/bmj.k601 PMID: 30002074

10. Corbin LJ, Richmond RC, Wade KH, Burgess S, Bowden J, Smith GD, et al. BMI as a modifiable risk factor for type 2 diabetes: Refining and understanding causal estimates using mendelian randomization. Diabetes. 2016; 65:3002–3007. https://doi.org/10.2337/db16-0418 PMID: 27402723

11. Holmes M-V, Asselbergs FW, Palmer TM, Drenos F, Lanktree MB, Nelson CP, et al. Mendelian randomization of blood lipids for coronary heart disease. Eur Heart J. 2015; 36:539–550. https://doi.org/10.1093/eurheartj/eht571 PMID: 24474739

12. FinnGen. FinnGen Documentation of R3 release. 2020 [cited 26 Oct 2020]. Available from: https://finngen.gitbook.io/documentation/.

13. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK Biobank resource with deep phenotyping and genomic data. Nature. 2018; 562:203–209. https://doi.org/10.1038/s41586-018-0579-z PMID: 30305743

14. Alser O, Craig RS, Lane JCE, Uribe AP, Robinson DE, Rees JL, et al. Serious complications and risk of re—operation after Dupuytren's disease surgery: a population—based cohort study of 121, 488 patients in England. Sci Rep. 2020;1–10. https://doi.org/10.1038/s41598-019-56847-4 PMID: 31913322

15. Sharp SA, Rich SS, Wood AR, Jones SE, Beaumont RN, Harrison JW, et al. Development and Standardization of an Improved Type 1 Diabetes Genetic Risk Score for Use in Newborn Screening and Incident Diagnosis. 2019; 42:200–207. https://doi.org/10.2337/dc18-1785 PMID: 30655379

16. Tyrrell JS, Yaghootkar H, Freathy RM, Hattersley AT, Frayling TM. Parental diabetes and birthweight in 236 030 individuals in the UK Biobank Study. 2013;1714–1723. https://doi.org/10.1093/ije/dyt220 PMID: 24336895

17. Loh P-R, Kichaev G, Gazal S, Schoech AP, Price AL. Mixed-model association for biobank-scale datasets. Nature genetics. 2018. pp. 906–908. https://doi.org/10.1038/s41588-018-0144-6 PMID: 29892013

18. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, et al. LocusZoom: regional visualization of genome-wide association scan results. 2010; 26:2336–2337. https://doi.org/10.1093/bioinformatics/btq419 PMID: 20634204

19. The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science. 2020; 369:1318–1330. https://doi.org/10.1126/science.aaz1776 PMID: 32913098

20. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics. 2010/07/08. 2010; 26:2190–2191. https://doi.org/10.1093/bioinformatics/btq340 PMID: 20616382

21. Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen W-M. Robust relationship inference in genome-wide association studies. Bioinformatics. 2010/10/05. 2010; 26:2867–2873. https://doi.org/10.1093/bioinformatics/btq559 PMID: 20926424

22. Bowden J, Smith GD, Burgess S. Mendelian Randomization Methodology Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. 2015;512–525. https://doi.org/10.1093/ije/dyv080 PMID: 26050253

23. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. Genet Epidemiol. 2016/04/07. 2016; 40:304–314. https://doi.org/10.1002/gepi.21965 PMID: 27061298

24. Oram RA, Patel K, Hill A, Shields B, McDonald TJ, Jones A, et al. A type 1 diabetes genetic risk score can aid discrimination between type 1 and type 2 diabetes in young adults. Diabetes Care. 2016; 39:337–344. https://doi.org/10.2337/dc15-1111 PMID: 26577414

25. Mahajan A, Go MJ, Zhang W, Below JE, Gaulton KJ, Ferreira T, et al. Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. Nat Genet. 2014; 46:234–244. https://doi.org/10.1038/ng.2897 PMID: 24509480

26. Morris AP, Voight BF, Teslovich TM, Ferreira T, Segrè A-V, Steinthorsdottir V, et al. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. Nat Genet. 2012; 44:981–990. https://doi.org/10.1038/ng.2383 PMID: 22885922

27. Howe LJ, Tudball M, Smith GD, Davies NM. Interpreting mendelian randomization estimates of the effects of categorical exposures such as disease status and educational attainment. medRxiv. 2020. https://doi.org/10.1101/2020.12.14.20248168

28. Kamal N, McGee SL, Eng K, Brown G, Beattie S, Collier F, et al. Transcriptomic analysis of adhesive capsulitis of the shoulder. J Orthop Res Off Publ Orthop Res Soc. 2020; 38:2280–2289. https://doi.org/10.1002/jor.24686 PMID: 32270543

29. Ng M, Thakkar D, Southam L, Werker P, Ophoff R, Becker K, et al. A Genome-wide Association Study of Dupuytren Disease Reveals 17 Additional Variants Implicated in Fibrosis. Am J Hum Genet. 2017; 101:417–427. https://doi.org/10.1016/j.ajhg.2017.08.006 PMID: 28886342

30. ten Dam E-JPM, van Beuge MM, Bank RA, Werker PMN. Further evidence of the involvement of the Wnt signaling pathway in Dupuytren's disease. J Cell Commun Signal. 2016; 10:33–40. https://doi.org/10.1007/s12079-015-0312-8 PMID: 26635199

31. Green HD. Manhattan Plots for visualisation of GWAS results. A function for plotting a Manhattan Plot in MATLAB directly from a text file of GWAS statistics from PLINK, BOLT-LMM or SAIGE. GB. 2019. Available from: https://uk.mathworks.com/matlabcentral/fileexchange/69549-manhattan-plots-for-visualisation-of-gwas-results.