Article

A genome-wide association meta-analysis implicates Hedgehog and Notch signaling in Dupuytren's disease

Sophie A. Riesmeijer [1,2] ✉, Zoha Kamali [2,3], Michael Ng [4], Dmitriy Drichel[5,6], Bram Piersma[7], Kerstin Becker[5], Thomas B. Layton[8], Jagdeep Nanchahal [8], Michael Nothnagel [5,6], Ahmad Vaez [2,3], Hans Christian Hennies [6,9], Paul M. N. Werker[1], Dominic Furniss [4] & Ilja M. Nolte [2]

Dupuytren's disease (DD) is a highly heritable fibrotic disorder of the hand with incompletely understood etiology. A number of genetic loci, including Wnt signaling members, have been previously identified. Our overall aim was to identify novel genetic loci, to prioritize genes within the loci for functional studies, and to assess genetic correlation with associated disorders. We performed a meta-analysis of six DD genome-wide association studies from three European countries and extensive bioinformatic follow-up analyses. Leveraging 11,320 cases and 47,023 controls, we identified 85 genome-wide significant single nucleotide polymorphisms in 56 loci, of which 11 were novel, explaining 13.3–38.1% of disease variance. Gene prioritization implicated the Hedgehog and Notch signaling pathways. We also identified a significant genetic correlation with frozen shoulder. The pathways identified highlight the potential for new therapeutic targets and provide a basis for additional mechanistic studies for a common disorder that can severely impact hand function.

Fibrosis, the excessive accumulation of extracellular matrix components, can affect nearly every tissue in the body and is increasingly recognized as a major cause of morbidity and mortality[1]. Dupuytren's disease (DD) is a fibrotic disorder of the fascias of the hand that causes fingers to irreversibly contract. It is also associated with excess mortality due to a wide range of causes, including cancer[2]. DD is very common and its prevalence increases with age, affecting 12% of Western population aged 55 years to 29% of those aged 75 years[3]. The initial presentation is as a highly cellular nodule that progresses to form fibrous cords. As the cords extend into the finger and undergo remodeling and shortening, patients develop contractures that impair function and quality of life[4]. Treatment of DD is currently limited to late-stage disease to reduce flexion contractures by disrupting or excising the cords. However, recurrence rates are high, and there is no cure for this debilitating disease[5].

Current understanding of the etiology of DD is limited, with a complex interaction between genetic and environmental factors[6]. There is strong evidence for an association with advanced age, male sex, family history of DD, heavy alcohol consumption, cigarette smoking, and manual work exposure[7]. Furthermore DD shares a genetic etiology with body mass index (BMI), type II diabetes mellitus (T2D), and levels of triglycerides and high-density lipoprotein (HDL)[8].

[1]University of Groningen, University Medical Center Groningen, Department of Plastic Surgery, Groningen, The Netherlands. [2]University of Groningen, University Medical Center Groningen, Department of Epidemiology, Groningen, The Netherlands. [3]Department of bioinformatics, School of Advanced Medical Technologies, Isfahan University of Medical Sciences, Isfahan, Iran. [4]Nuffield Department of Orthopaedics, Rheumatology, and Musculoskeletal Science, Botnar Research Centre, University of Oxford, Oxford, UK. [5]Cologne Center for Genomics, University of Cologne, Cologne, Germany. [6]Faculty of Medicine and the Cologne University Hospital, Cologne, Germany. [7]University of Groningen, Groningen, The Netherlands. [8]Kennedy Institute, University of Oxford, Oxford, UK. [9]Department of Biological Sciences, University of Huddersfield, Huddersfield, UK. ✉e-mail: sophieriesmeijer@gmail.com

Intriguingly, adiposity is causally protective against DD[9,10]. Genome-wide associations studies (GWASs) have so far identified 61 genetic risk variants[11–13]. The overall (broad-sense) heritability of DD was estimated to be 80%, of which 67% is attributable to common genetic variants[8,14]. However, the variance explained by known genetic variants was estimated as 11.3%[12]. Thus, much of the genetic susceptibility for DD remains to be elucidated.

Development of novel therapeutic approaches requires a systematic approach of identification of novel genetic risk factors, broadening our understanding of the mechanisms involved with these genetic susceptibility loci, and studying their functional consequences. Here we addressed these aims by performing meta-analysis of GWAS data from six cohorts to prioritize genes involved in DD.

Results

Study cohorts
Sample and genotyping details of the cohorts included in the meta-GWAS are provided in Supplementary Data 1. An overview of the QC output parameters of the GWAS per cohort is given in Supplementary Data 2.

Association analysis
The meta-analysis included a total of 11,320 cases and 47,023 controls, tested at 8,123,121 variants after QC. Clumping analysis identified 85 independent genome-wide significantly associated SNPs in 56 loci (Fig. 1; Supplementary Data 3; Supplementary Fig. 2). Forty-five of these loci have been previously reported[12,15], 11 represent new loci. In addition, we identified 24 new secondary hits in 12 known loci. All 26 previously identified loci from the previous (UK) GWAS[12] were confirmed, but for six loci the effects were found to be heterogeneous between the six cohorts and were therefore excluded from our meta-GWAS (Supplementary Data 4). We also found associations with 45 of the 61 loci implicated by Agren et al.[13]. Forest plots of the significantly associated SNPs are shown in Supplementary Fig. 3. Thirty-four of the 85 genome-wide significant meta-GWAS SNPs were replicated in the FinnGen cohort and 77 reached genome-wide significance in the combined meta-analysis with FinnGen (Supplementary Data 5). Five of the 85 genome-wide significant SNPs were not available in FinnGen. Gene-based analysis is shown in Fig. 2.

Bioinformatic follow-up analyses
In silico annotation. In silico sequencing analysis returned 4542 and 2111 variants in moderate ($r^2 > 0.5$) and high ($r^2 > 0.8$) LD, respectively, with the identified 85 meta-GWAS SNPs (Supplementary Data 6). One of the meta-GWAS SNPs (rs1042704) as well as eight correlated SNPs were non-synonymous. Out of these eight nsSNPs, three were in high LD ($r^2 > 0.8$) with two meta-GWAS SNPs: rs366905 and rs34412930. The nine nsSNPs mapped to the genes *TMEM81*, *DSTYK*, *SUMO4*, *CFTR*, *TNC*, *MMP14*, and *LDHAL6B* (three nsSNPs). We flagged four nsSNPs in *TMEM81*, *DSTYK*, *MMP14*, and *LDHAL6B* as deleterious based on scaled CADD scores >20. This score means that these missense mutations are among the top 1% of all possible substitutions in the human reference genome (~8.6 billion) ranked by deleteriousness based on over 60 different annotation sources[16]. This score is also larger than the median for all non-synonymous variations (i.e. 15).

In silico pleiotropy analysis (lookups of associated phenotypes) of all meta-GWAS SNPs and their correlated SNPs revealed associations with hematologic and certain anthropometric traits (Supplementary Data 6). Interestingly, two SNPs were associated with keloid, another fibroproliferative disorder influenced by transforming growth factor beta (TGF-β1) production and Wingless and Int-1 (Wnt) signaling[17]. Furthermore, rs11581010 has been previously associated with lower BMI[18]. Additional associations were found with variance of red cell counts (red cell distribution width), heel bone mineral density, and type II diabetes mellitus. The locus harboring *ZBTB40* was the most pleiotropic region, with reported GWAS associations with a variety of hematologic traits, bone mineral density traits, and chronic inflammatory diseases.

Gene prioritization, pathway, and tissue prioritization analyses. The full gene prioritization analyses can be found in the Supplementary Results. Since effect sizes of blood and fibroblast eQTLs showed a high similarity ($r = 0.71$; Supplementary Fig. 4), blood eQTL data were used, given the much larger available sample size. A list of 119 prioritized genes resulted from these analyses (Supplementary Data 7). These genes were then assessed for association of differential gene expression in fibroblasts and further functional analyses. The FINEMAP analysis showed that most of the 56 loci the index SNPs had considerable evidence for being causal themselves and that the proportion of variants with a considerable evidence was the highest for exonic variants (see Supplementary Data 8 and Supplementary Fig. 10).

Functional enrichment analysis showed that our results were enriched for genes involved in abnormal limb, cartilage, and skeleton morphology (Supplementary Data 9). Further functional analysis suggested deeper mechanistic insights, most importantly TGF-β signaling, epithelial cell migration and cell-matrix adhesion (Supplementary Data 10). Protein-protein interaction network of the prioritized genes revealed two connected components, of which the first one was suggestive of response to stress and the second one was suggestive of viral/bacterial infection (Supplementary Data 24).

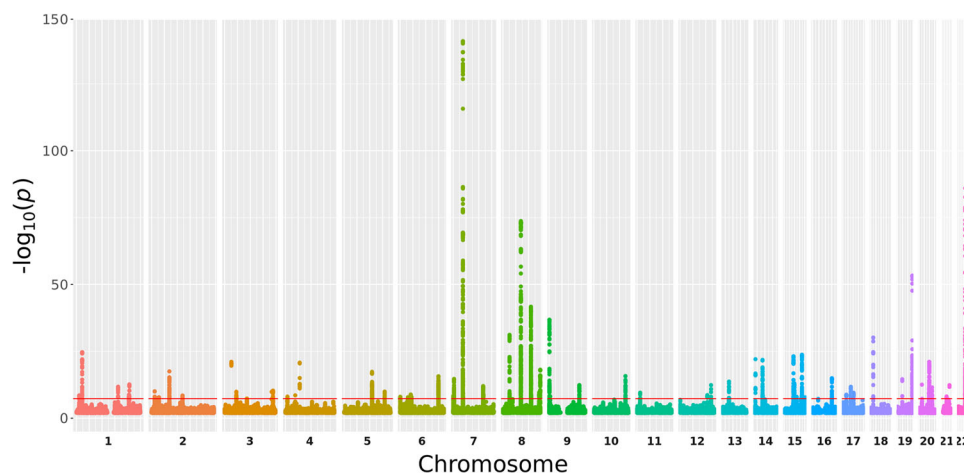

Fig. 1 | Manhattan plot of the logistic regression for the meta-GWAS association analysis. The horizontal red line represents a *p*-value threshold of 5×10^{-8} (e.g. the multiple comparison correction for genome-wide significance).
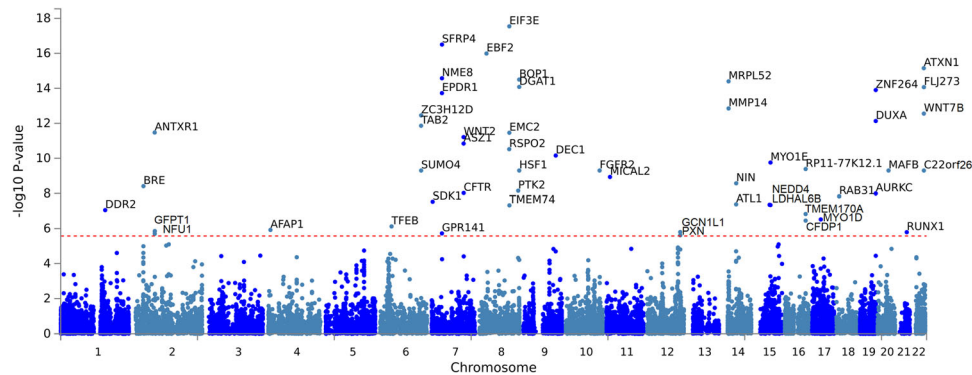
Fig. 2 | Gene-based analysis as computed by MAGMA, using a multiple regression model. Meta-GWAS summary statistics were mapped to 18,879 protein coding genes. Genome-wide significance (red dashed line in the plot) was defined by Bonferroni correction at $p = 0.05/18,879 = 2.65 \times 10^{-6}$.
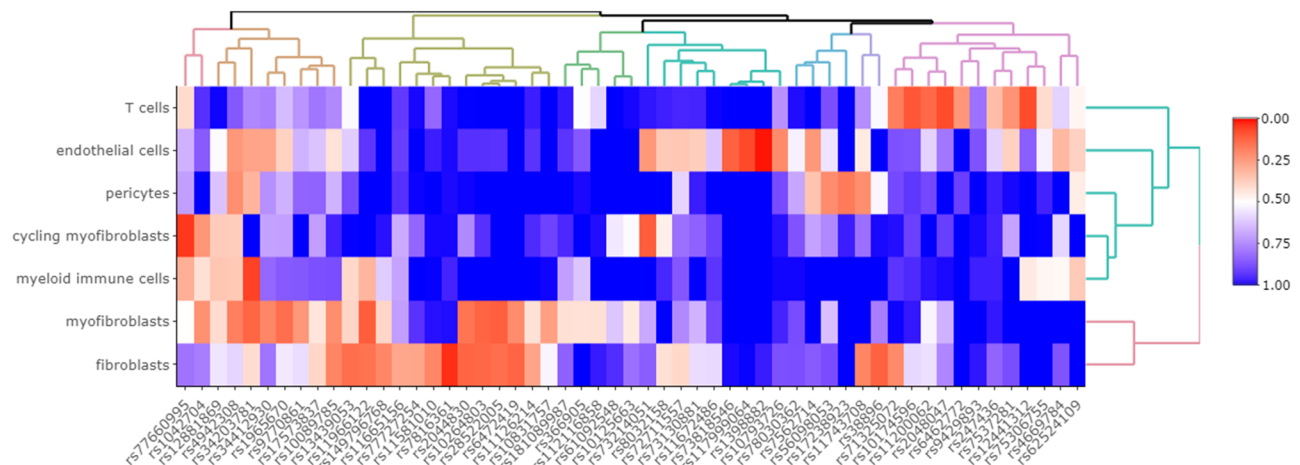

Fig. 3 | Cell population-relevant genes. Heatmap showing specificity scores (between 0 [red] and 1 [blue] where a lower value indicates greater specificity to the cell type), produced by SNPsea, of meta-GWAS associated DD loci in cell populations derived from single cell RNA-seq of DD nodules ($n = 6$ patients)[60]. Dendrograms (clustering trees) can be observed for cell populations (y-axis) as well as genetic loci (x-axis).

Tissue prioritization results showed enrichment of 12 tissues for the expression of DD prioritized genes (FDR < 0.01; Supplementary Data 11), on top of which fibroblasts showed the highest expression levels (Supplementary Fig. 5).

Transcriptome-wide association study. SMR analysis with fibroblast eQTLs revealed seven genes with expression levels significantly associated with DD ($P_{SMR} < 6.84 \times 10^{-6}$)[19]. Four of these associations remained significant after the heterogeneity test ($P_{HEIDI} \geq 7.14 \times 10^{-3}$, Supplementary Data 12). These include *RPLP0*, with a higher risk for DD, and *CTD-2587M2.1*, *DLG5*, and *TEAD3* with predicted protective effects against DD. The long non-coding RNA *CTD-2587M2.1* was associated in SMR analysis with blood eQTLs, with consistent direction of effect (Supplementary Data 13).

Of the 119 prioritized genes for DD, 14 reached Bonferroni-corrected significance in the fibroblast SMR analysis (Supplementary Data 7). Five of these (*TNC, AFAP1, CHSY1, NEDD4,* and *CFDP1*) were identified in at least three bioinformatic follow-up analyses and fibroblast SMR analysis.

Cell population-relevant genes. Next, we aimed to identify cell types whose function is likely to be influenced by genes in our risk loci. Of the 85 meta-GWAS hits, SNPsea was unable to identify SNPs rs12442366 and rs886423. Twelve loci did not contain any genes. SNPsea merges SNPs with shared genes into single loci to avoid multiple counting of genes, thus resulting in a dataset of 55 gene sets. Myofibroblasts showed the strongest association to the meta-GWAS risk loci ($p = 0.08$). Moreover, automatic clustering highlighted that myofibroblasts and fibroblasts have more similar expression of genes in our meta-GWAS loci than other cell types (Fig. 3). In Supplementary Data 14 genes with the greatest specificity to each cell population are given for each combination of SNP and cell type.

Polygenic risk scores
The distributions of standardized PRS were in highly significantly different between cases and controls for both Dutch cohorts (CytoSNP OR = 3.25, 95% CI = 2.92-3.62, $p = 2.0 \times 10^{-102}$; GSA OR = 3.16, 95% CI = 2.87-3.48, $p = 3.9 \times 10^{-122}$; Fig. 4) and for both UK cohorts (UK Biobank OR = 2.94, 95% CI = 2.82-3.07, $p = 1.0 \times 10^{-310}$; BSSH-GODD OR = 3.48, 95% CI = 3.26-3.72, $p = 1.9 \times 10^{-307}$; Fig. 4) Liability-adjusted Nagelkerke's pseudo R^2 measures showed that 13.3% of the variance can be explained by the PRS in the Dutch CytoSNP cohort, 15.3% in the Dutch GSA cohort, 30.0% in the UK Biobank cohort, and 38.1% in the BSSH-GODD cohort. Very recently the Dutch Lifelines cohort was further genotyped using another array (FinnGen Thermo Fisher Axiom® custom array) for 28,500 additional participants, of whom 110 reported to have DD by questionnaire and 6,148 unrelated individuals were selected as controls. For this cohort, the PRS was less significant (OR = 2.44, 95% CI = 1.74-3.43, $p = 2.3 \times 10^{-7}$) and the liability-adjusted Nagelkerke's pseudo R^2 was 13.3%.
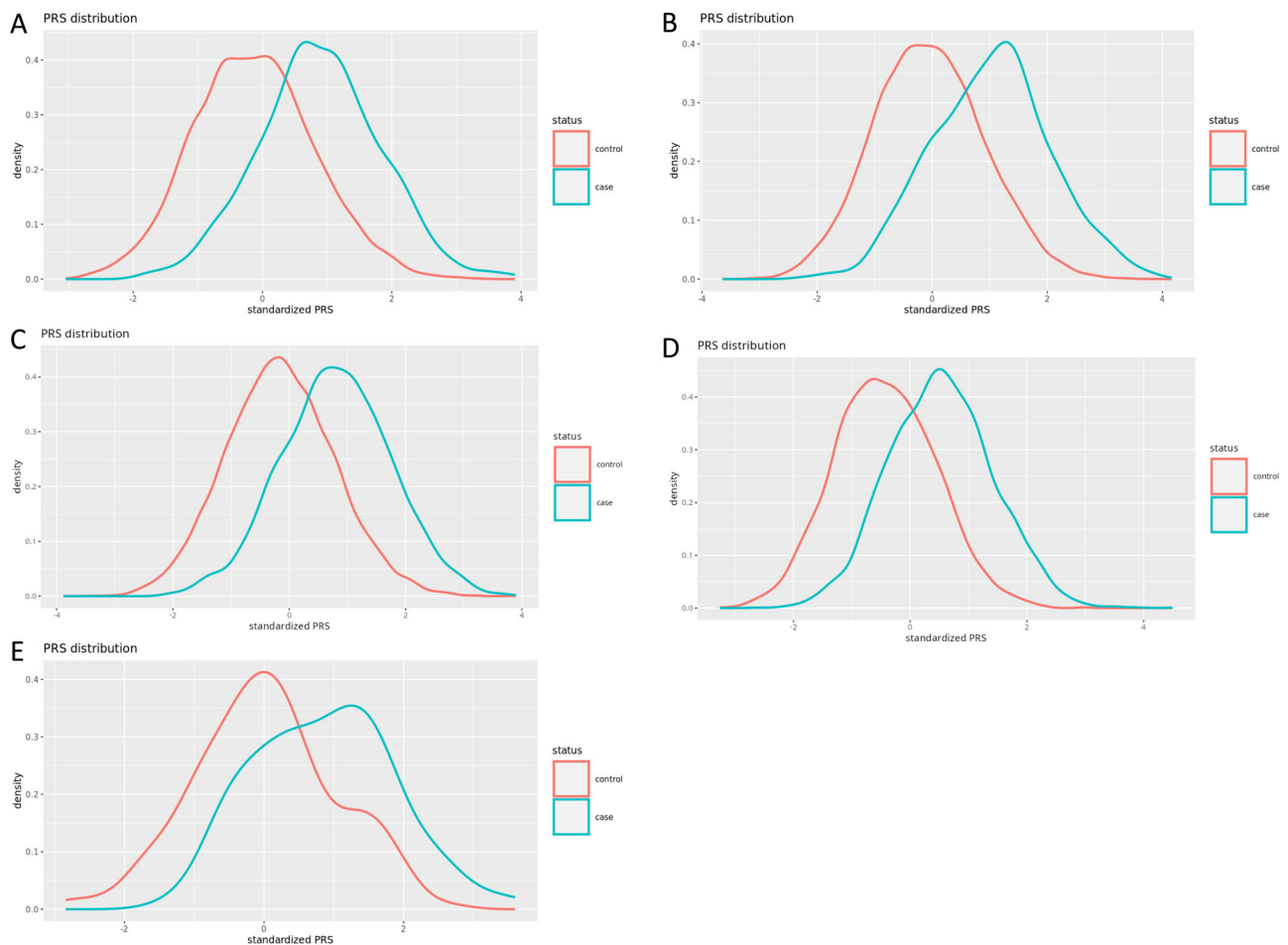
Fig. 4 | Variance explained through polygenic risk score analyses. Density plot of the PRS distribution for cases and controls in (**A**) the Dutch CytoSNP cohort, (**B**) the Dutch GSA cohort, (**C**) UK Biobank cohort (**D**) the UK BSSH-GODD cohort, and (**E**) the Dutch FinnGen cohort.

Genetic correlations

Frozen shoulder showed a significant positive genetic correlation of 0.30 with DD (p-value $= 1.9 \times 10^{-6}$), suggesting that there is an overlap between causal variants for DD and for frozen shoulder (Supplementary Data 15). Increasing BMI was significantly negatively correlated to DD (p-value $= 1.2 \times 10^{-8}$, $r = -0.14$), and increasing HDL was positively correlated (p-value $= 3.6 \times 10^{-5}$, $r = 0.12$). Genetic correlations for fasting glucose, HbA1c, triglycerides, idiopathic pulmonary fibrosis, psoriasis, systemic sclerosis, T2D, bone mineral density, and height with DD were not significant.

Colocalization analysis

The significantly genetically correlated traits frozen shoulder, BMI, and HDL were considered for colocalization analysis. For frozen shoulder, five genome-wide significant SNPs have been identified[20]. Colocalization analysis revealed that DD and frozen shoulder share the same causal variants rs1042704 (posterior probability 0.99) and rs28606049 (posterior probability 0.99). SNP rs2472660, which was also located in a locus associated to DD, was not found to be causally related to both disorders. For BMI, two previously associated loci were within a locus also associated with DD: rs10779751 and rs7607490[21]. For rs10779751, neighboring SNP rs11581010 was revealed as a shared causal variant for both DD and BMI (posterior probability 0.99). For rs7607490, none of the SNPs within its region had a causal effect on both BMI and DD (posterior probability 0.12). None of the previously associated HDL SNPs were at the same loci as genome-wide significant DD SNPs, thus colocalization analysis for HDL could not be performed.

Other gene prioritization analyses

Results from supplementary methods and results are available in the supplements: list of prioritized genes, Supplementary Data 17; co-regulation analysis, Supplementary Data 18; multi-layer analysis, Supplementary Data 19; multi-QTL analysis, Supplementary Data 20; functional enrichment analysis of 73 prioritized genes, Supplementary Data 21; functional enrichment analysis of 23 prioritized genes, Supplementary Data 22; functional enrichment analysis of 119 prioritized genes, Supplementary Data 23 Protein-protein interaction (composite) network analysis, Supplementary Data 24.

Discussion

Although DD is a common disease with a strong genetic component, underlying genetic factors and disease mechanisms are widely elusive. Therefore, the aims of this study were to identify additional genetic loci and pathways for DD to further explain the genetic variance of DD, and to provide directions for future functional follow-up studies. We performed the largest GWAS of DD to date, meta-analyzing six cohorts, including 11,320 cases and 47,023 controls. It yielded 85 genome-wide significantly associated SNPs at 56 loci, 34 of which have not been previously described. With these variants thrice the amount of phenotypic variance (narrow-sense heritability) can be explained, up to 38.1% in contrast to 11.3%[12]. Nevertheless, a large amount remains undiscovered considering the estimated heritability of 80%[8,12,14].

Replication of our genome-wide significant SNPs in the FinnGen cohort revealed only a moderate number of replicated SNPs (34 out of 85). This moderate overlap of SNPs available from both datasets might be explained by the fact that the Finnish population is known to have

experienced a population bottleneck and therefore to be genetically diverse from other European populations[22]. Consequently we chose to report all 85 SNPs as DD associated variants and not only those that could be replicated in FinnGen. In addition, FinnGen used a Thermo-Fisher Axiom custom array containing variants enriched in Finland, thereby explaining why many of the suggestive associations found by FinnGen were not available in our meta-GWAS. Nevertheless, the majority ($n = 77$) of the 85 genome-wide significant SNPs from our meta-GWAS reached a genome-wide significance in the meta-analysis of our meta-GWAS results and those from the FinnGen cohort.

Very recently, a meta-analysis of DD GWAS was performed using the results from three biobanks (UK Biobank, FinnGen, and Michigan Genomics Initiative)[13]. For this study, the authors identified DD cases based on available International Classification of Diseases codes (ICD, Ninth Revision, Clinical Modification; code 728.71) available from hospital inpatient data. In the current study, we used DD cases diagnosed and/or treated by a plastic or hand surgeon in our outpatient clinics. Due to the difference in diagnosing, these cohorts are likely not quite comparable regarding phenotyping, in addition to the above-mentioned genetic diversity of the studied populations. We thus chose not to include these cohorts in our meta-analysis, but rather compare the results. From the 56 regions that we identified, Agren et al. also implicated 45, among which 25 that were already reported previously[11–13]. The eleven additional loci include the genes *MTOR, BABAM2, LRRC3B, LIMD1, LPP, AFAP1, LOC101927691, TFEB, MICAL2, MEOX1*, and *ADAMTS5*.

To prioritize genes that substantially impact DD phenotype we performed extensive bioinformatic follow-up analyses of our 85 genome-wide significantly associated SNPs in 56 loci in order to identify the most likely causal variants within these loci. These analyses identified 40 functional SNPs on the basis of it being a non-synonymous SNP, an eQTL, a multiQTL, or being colocalized with DD-related phenotypes (Supplementary Data 17. Although prioritized genes require experimental validation to reveal the mechanistic link with associated SNPs, they implicated the Hedgehog (Hh) and Notch signaling pathways. While Notch signaling has very recently been associated with DD[23], we newly report on Hh signaling. Both may contain potential therapeutic drug targets. In addition, we also found further candidates in the Wnt/β-catenin and Hippo signaling pathways, which are already known to be involved in the pathogenesis of DD[11,12,24–26]. Therefore, these signaling pathways and their prioritized genes are discussed in more detail below and visualized schematically in Supplementary Fig. 6. We focus our discussion on genes identified in at least three gene prioritization analyses and replicated in the fibroblast SMR analysis (Supplementary Data 16), as we consider these results to be robust.

Hh signaling plays a crucial role in embryonic development, regulating differentiation, proliferation, and tissue patterning of the brain, internal organs, and limbs[27]. The importance of Hh signaling in fibrosis has already been established in both animal models and humans[27]. In humans, Hh signaling is implicated in fibrotic kidney disease, pancreatic fibrosis, liver fibrosis, and biliary fibrosis[28–30]. Hh signaling can be activated by injury[28–30]. Profibrotic factors such as TGF-β1 can activate expression of Hh members of the GLI family in human fibroblasts[31]. CHSY1 is a regulator of Hh signaling. It encodes an enzyme that plays a critical role in the biosynthesis of chondroitin sulfate, a glycosaminoglycan involved in many biological processes including cell proliferation and morphogenesis. To our knowledge we are the first to report a link between CHSY1 and fibrosis (see Supplementary Fig. 6). Future research on the role of *CHSY1* in DD should focus on studying its expression levels in DD myofibroblasts. The extracellular matrix (ECM) glycoprotein gene *TNC* (encoding tenascin C) is a target of the Hh signaling pathway transcription factors GLI1 and GLI2 (Supplementary Fig. 6)[27]. Berndt et al. found that tenascin C was one of the constituents of the extracellular matrix formed by Dupuytren's myofibroblasts[32]. In kidney fibrosis tenascin C was

induced by Sonic Hedgehog (SHH) and identified as a major constituent of promotion of fibroblast proliferation[33]. In our analyses tenascin C was found to be involved in one of the two major connected components of DD protein-protein interaction networks, which was enriched in ECM organization. As we also identified a nsSNP in *TNC*, functional studies might be particularly fruitful. DLG5 is a member of the MAGUK superfamily that is involved in maintenance of epithelial cell polarity, cell proliferation control, and cell migration and invasion[34]. In mice Dlg5 is required to interact with Hh receptor Smoothened (Smo) for Gli protein activation[35]. In mouse fibroblasts lacking Dlg5, Hh-induced Smo accumulation was observed[35].

Several drugs targeting the Hh signaling pathway are being or have been developed, studied, and manufactured. Cyclopamine, a small molecule inhibitor of the transmembrane protein SMO, attenuated renal fibrosis in vivo[27]. Hh pathway inhibitors have been studied as cancer drugs[36–38]. Vismodegib also inhibits SMO, blocking activation of GLI proteins to transcribe Hh target genes. Thus, cyclopamine and vismodegib might have mitigating effects on the DD phenotype.

Notch signaling is a highly conserved embryonic developmental signaling pathway. It consists of several receptors, NOTCH1–NOTCH4, and their ligands, delta-like and jagged (JAG). Activation of the pathway usually occurs via expression of the ligand in a signal-giving cell[39]. Notch signaling has been shown to be activated in retinal, renal, and hepatic fibrosis[40–43]. In vessels of the microcirculation in DD nodules, pericytes (that surround the endothelial cells) were shown to specifically express NOTCH3[23]. Furthermore, the γ-secretase inhibitor XX (GSIXX, referred to as DBZ), a pharmacological inhibitor of Notch, effectively ameliorated renal fibrosis in mice[43]. *CHSY1* is a member of the Fringe family of genes that modulate Notch signaling via ligand interaction with Notch receptors[44]. Tian et al. described overproduction of JAG1 and subsequent Notch activation in absence of CHSY1 in fibroblasts from patients with syndromic brachydactyly associated with a truncating frameshift mutation in *CHSY1*[45]. Knockdown of CHSY1 promotes Notch signaling, and overexpression of CHSY1 reversed Notch activation[45], consistent with the predicted protective effect on DD that we found for *CHSY1* in the SMR analysis with fibroblast data.

The Hippo signaling pathway has been previously implicated in fibrosis and DD[25,26]. We found a new association with *TEAD3*, a member of the TEA domain family of transcription factors that are essential in mediating YAP-dependent gene expression. YAP1 is a regulator of myofibroblast differentiation and contributes to the maintenance of the contractile phenotype in DD myofibroblasts[25]. As TEAD3 is a key transcription factor mediating YAP function[46], we hypothesize that its decreased expression might up-regulate YAP dependent gene expression in DD.

We have demonstrated that the previously observed phenotypic association of frozen shoulder with DD[47] likely results from a substantial genetic correlation. Moreover, we showed that frozen shoulder and DD share two causal variants among the three variants that are shared between the two diseases in total. Of these, rs1042704 is an exonic SNP in matrix metalloproteinase 14 gene (*MMP14*), that has been shown to cause a specific defect in collagenolytic activity in DD derived fibroblasts[48]. Interestingly, in a series of 12 people treated for an inoperable gastric carcinoma with a synthetic matrix metalloproteinase inhibitor, half developed frozen shoulder or a condition resembling DD[49]. These findings further underline the importance of MMPs in both frozen shoulder and DD[48]. The second shared causal variant rs28606049 is an intron variant in *WNT7B*, which is highly upregulated in DD[24,50]. As inferred from the substantial genetic correlation, many more variants are likely shared between frozen shoulder and DD. We also reproduced previously described genetic correlations with BMI and HDL, but not for triglycerides and T2D. BMI and DD only shared one causal variant, indicating that LD between associated SNPs or (mediated) pleiotropic effects of non-genome-wide significant SNPs

are likely the driving force behind the genetic correlation[51]. This is further underlined by our in silico pleiotropy analysis, where a substantial enrichment of DD loci in a previously conducted BMI GWAS was found. The same theory holds true for HDL and DD, as none of the genome-wide significant SNPs for HDL resided at the same loci as those for DD.

Strengths of this study include the large sample size of DD patients and controls and the thorough phenotyping of the DD patients by a plastic or hand surgeon, resulting in the identification of many additional loci. Furthermore, the inclusion of cohorts from multiple European countries achieved translatability of results across multiple European populations (except for Finns, as discussed above), instead of only one subpopulation. Moreover, genotype imputation facilitated integration of genotype data from multiple cohorts analyzed with different arrays and enhanced the power for detecting SNPs. We performed a multitude of bioinformatic follow-up analyses from different mechanistic angles and only took genes into account that were prioritized in at least three analyses and replicated in a tissue specific analysis. Therefore, we argue that the results are robust. Although we provided starting points for functional studies, a limitation of this study is that we did not investigate experimentally the effects of associated SNPs and prioritized genes in vitro. Unfortunately age and sex were not available for some cohorts. Therefore in these cohorts cases and controls could not be age and sex matched and these parameters could not be included in the analysis as covariates. However the result is likely that some statistical power is lost in case the controls were younger than the cases, as the control cohort may still include individuals who became cases after inclusion in the study. In case controls were older than cases, this could have introduced bias due to increased mortality, however this is likely not of great influence in the case of DD. In addition we excluded heterogeneous effects from the meta-analysis results, which will have filtered out a spurious result due to potential age or sex mismatching in one of the cohorts. Another limitation of this work is that we were not able to validate our meta-GWAS results in a genetically similar cohort, as the replication analysis used the FinnGen population which is genetically diverse from other European populations[22]. The generalizability of our results, acquired studying populations of European ancestry, to other ethnic populations, is likely limited. To our knowledge, no GWAS have been performed in populations other than from European ancestry. Increasing diversity of ancestries among GWAS study participants can advance our understanding of the genetic susceptibility to DD for all populations. We believe a next step in genetic epidemiological research into DD would be performing multi-ancestry GWAS, perhaps through utilizing multi-ethnic biobanking studies due to the lack of available cohorts.

In conclusion, this meta-analysis of six GWASs identified 34 novel loci for DD and newly implicated the Hh signaling pathway and confirmed association of the Notch signaling pathway in the etiology of DD. Prioritized genes *CHSY1*, *NEDD4*, and *DLG5* have regulatory properties in Notch, Hh, Hippo, and Wnt signaling. These pathways contain therapeutic targets for which a number of inhibitors exist. We have outlined starting points for future mechanistic studies for DD. Additionally, we found a genetic correlation between frozen shoulder and DD and identified two SNPs that are causal variants for both frozen shoulder and DD. Our data will help inform future mechanistic studies to validate therapeutic targets and develop new treatment strategies for DD. Until then, the increased knowledge about the genetic susceptibility to DD provided by this meta-GWAS facilitates research into individualized risk prediction for DD through genetic profiling.

Methods
Study cohorts
We used data from six cohorts with DD cases and healthy controls from the Netherlands, United Kingdom (UK), and Germany. The cases

were individuals of European ancestry who had been diagnosed with DD by a plastic or hand surgeon and/or who had undergone surgical treatment for DD. Controls were population-based subjects from the Lifelines cohort study (the Netherlands), from the UK Household Longitudinal Study (UK), the UK Biobank initiative (UK), and from the PopGen and KORA studies (Germany) with no known diagnosis of DD. All study populations were described in detail previously[11,12,52–59]. All samples analyzed in previous GWASs[11,12,51,54,60] were included in these cohorts.

Ethical approval
The studies used in this meta-analysis were approved by the Research Ethics Committee or equivalent at all institutions where the data were collected: (1) The Genetic Origin of Dupuytren Disease (GODDAF) Study (the Netherlands) was approved by the Ethics Committee of the University Medical Center Groningen, document number 2007/067; (2) The Lifelines study (the Netherlands) was approved by the Ethics Committee of the University Medical Center Groningen, document number 2007/152. This study ('The role of genetic variants in Dupuytren disease') has Lifelines study ID OV18_0461; (3) The British Society for Surgery of the Hand Genetics of Dupuytren's Disease (BSSH-GODD) study (United Kingdom) was approved by the Oxfordshire Research Ethics Committee, document number B/09/H0605/65; (4) The UK Biobank (United Kingdom) was approved by the North West Multi-Centre Research Ethics Committee, document number 11/NW/0382. This study ('The Genetics and Epidemiology of Common Hand Conditions') has UK Biobank study ID 22572; (5) The German Dupuytren Study was approved by the Ethics Commission of the Faculty of Medicine of the University of Cologne, document number 14/292. The KORA study was approved by the Ethics Committee of the Bavarian Medical Association (Bayerische Landesärztekammer) and the Bavarian commissioner for data protection and privacy (Bayerischer Datenschutzbeauftragter). The PopGen study was approved by the Ethical committee of the Medical Faculty of Christian-Albrechts-Universität (CAU), Kiel. Informed consent was obtained from all subjects in accordance with Declaration of Helsinki protocols.

Genotyping, quality control and imputation procedures
The genotyping, quality control (QC), and imputation procedures of the UK and German cohorts were described in detail previously[12,51–53,60] A detailed pipeline of genotype QC and imputation procedures for the Dutch case cohorts can be found in the Supplementary Materials.

GWAS
GWAS for the Dutch cohorts were performed with logistic regression analysis in PLINK (version 1.9[61–64]. For each single nucleotide polymorphism (SNP) an analysis was performed with disease status as outcome and age, sex, and the first ten principal components as covariates. For the UK and German cohorts, for each SNP a logistic regression was performed with sex and principal components as covariates, calculated with PLINK (version 1.9) or SNPTEST (version 2.5.4-beta3), respectively[65,66]. Age was not available in these cohorts. The principal components were calculated using PLINK (version 1.9[61,64]. Quality control of the GWAS summary statistics was performed in R (version 3.6.1) using the GWASinspector package for each cohort separately[67]. In case of quality issues the respective cohort was notified and problems were solved. Using the QQ plots from GWASinspector, for each cohort specific imputation quality and allele frequency thresholds were set (see Supplementary Methods). Genomic reference build 37 (GRCh37/hg19) was used in this study.

Meta-analysis
Meta-analysis of the six GWAS results (meta-GWAS) was performed using METAL with fixed effects inverse variance weighting method[68]. We used a double genomic control correction to control for genomic

inflation due to population stratification within and between study cohorts[69]. After meta-analysis, only high-quality variants were considered for follow-up analyses: i.e. variants present in three or more (out of six) cohorts, variants present in >10,000 participants (~20% of total), and variants with a heterogeneity p-value > 0.05 (alongside other quality criteria) were considered for follow-up analyses. These filtered meta-GWAS summary statistics (containing only high-quality variants) were used for gene-based analysis through Functional Mapping and Annotation of Genome-Wide Association Studies (FUMA)[70].

We sought to replicate significant loci discovered in our meta-GWAS in the FinnGen cohort consisting of 3248 cases and 197,724 controls and performed a combined meta-analysis for these loci of our meta-GWAS and FinnGen's results[15]. For replication, SNPs with a one-sided p-value < 0.000588 (=0.05/85, i.e. Bonferroni correction) in FinnGen were considered replicated.

Identifying independent loci and secondary signals
SNPs were regarded as genome-wide significant if the p-value was $<5 \times 10^{-8}$. Genome-wide significant loci were considered independent when SNPs in neighboring loci were separated by >1 Mb on the same chromosome or on different chromosomes. If a locus contained only one significant SNP, it was considered unreliable and was discarded. To ascertain secondary signals, clumping was performed in PLINK ($r^2 \leq 0.01$; distance < 1 Mb) at each genome-wide significantly associated genetic locus[65].

Bioinformatic follow-up approach
In silico annotation. To uncover functional characteristics of the independent genome-wide significant SNPs and their surrounding regions, we determined which SNPs were in at least moderate linkage disequilibrium (LD) ($r^2 > 0.5$) with the GWAS SNPs based on 1000 G European reference data (with a maximum of 2 Mb distance) and annotated these using ANNOVAR (i.e. an in silico bioinformatics-based annotation approach)[71]. Variant call format (VCF) files for individuals of the European continental population from the 1000 Genomes Project, Phase 3 (version v5a, Feb. 20th 2015) were downloaded and data of the 2 Mb region surrounding each of the 85 identified SNPs were extracted with VCFtools[62,72]. The r^2 between identified SNPs and all biallelic SNPs residing in the 2 Mb region was calculated with PLINK[65]. SNPs were deemed sufficiently correlated if $r^2 > 0.5$. Identified and correlated SNPs were annotated using ANNOVAR (version October 2019) for functional consequences[73]. Nonsynonymous (ns, e.g. protein altering) SNPs were then characterized for their damaging effect using Combined Annotation Dependent Depletion (CADD) scores[16]. A scaled C-score > 20 was considered deleterious[16]. To better understand the possible function of the SNPs, we also performed an in silico pleiotropy analysis, that is, checking for association of all identified SNPs and SNPs in linkage disequilibrium (LD) ($r^2 > 0.5$) with other traits and diseases available in the GWAS Catalog database (version 21 April 2021)[73,74].

Gene prioritization, pathway, and tissue prioritization analyses. We followed up our GWAS results with a multi-omics post-GWAS approach[75,76] to gain insight on the biology of DD phenotype and identify potential key players in disease pathogenesis (Supplementary Fig. 1, Supplementary Methods), including FINEMAP to identify the most likely causal variants within each locus[77]. We used blood eQTL data for two gene prioritization analyses (as a discovery) and fibroblast eQTL data as a replication analysis of significant discovery findings, as proposed by Qi et al.[78]. We correlated blood and fibroblast eQTL data to check whether the use of blood data with large sample sizes for discovery, then fibroblast data for validation, was justified.

Cell population-relevant genes. Next, we estimated enrichment of genes implicated by genetic risk loci in DD cell populations, since

genes expressed in relatively few cell types are hypothesized to affect cell type functions. We studied single cell RNA sequence (scRNAseq) data from nodules of six DD patients containing seven cellular subtypes[60]. Using SNPsea, we prioritized cell types that are specific for Dupuytren's disease. We mapped the associated loci to candidate genes, calculated specificity scores of these genes to cell populations, and tested their significance[79]. We used default settings, except for '−slop' (1×10^5), '−min-observations' (100), and '−max-iterations' (1×10^6). A heatmap showing specificity scores of significant meta-GWAS loci in cell populations was created with the R-package heatmaply (version 1.3.0)[80]. For clustering we used heatmaply's default Euclidean distance measure and the average linkage function[80].

Polygenic risk score calculation
We calculated polygenic risk scores (PRS) for the two Dutch and two UK cohorts in order to study the variance explained (i.e. narrow-sense heritability) by genetic risk variants identified in the meta-GWAS. First, we re-ran the meta-analysis of GWASs with the same settings as described above, using a leave-one-out approach to acquire summary statistics that are independent of said cohort and could thus be used for PRS calculation. Then, we used SBayesRC (v.0.2.0) to improve polygenic risk prediction by integrating effect sizes from our summary statistics with functional genomic annotations[81]. Next, we constructed PRSs with PLINK using the improved weights calculated by SBayesRC[65]. Logistic regression analysis was performed to associate the PRS with the outcome DD adjusting for age, sex and principal components (PCs). Last, we calculated liability-adjusted Nagelkerke's pseudo R^2 measures to scale the phenotypic variance explained by PRS to the (previously estimated) disease prevalence of 7.08% (Dutch population) and 13.40% (UK population) that has been corrected for the general population of all ages, instead of the disease prevalence in the study population which was affected by ascertainment bias[82,83].

Genetic correlations
We assessed shared genetic etiology between DD and twelve clinically associated traits: body mass index (BMI), high-density lipoprotein (HDL), triglycerides, type 2 diabetes mellitus (T2D) adjusted for BMI, T2D unadjusted for BMI, T2D in European UK Biobank subjects with only variants available from the Haplotype Reference Consortium, fasting glucose, HbA1c, idiopathic pulmonary fibrosis, systemic sclerosis, frozen shoulder, psoriasis, bone mineral density, and height[20,21,55,64,84–93]. We employed LD score regression to calculate genetic correlations, using full GWAS summary statistics of DD and traits of interest[94]. We assessed the significance of the genetic correlation with a Bonferroni-corrected threshold being $\alpha = 0.05/14 = 0.0036$.

Colocalization analysis
For traits sharing a significant genetic correlation, we first selected genetic loci significantly associated to both traits. Then, SNPs from a 200 kb block surrounding the significantly associated SNPs were selected. Colocalization analysis was performed with the coloc and the Sum of Single Effects (SuSie) R-packages[95], using data from the 1000 Genomes Phase 3 dataset from the European continental population (version v5a, Feb. 20th 2015)[62], and calculated raw inter-variant allele count correlations (−r) with PLINK[65]. Missing correlation values were set to zero. Colocalization analysis estimates the posterior probability of a shared causal variant, a high posterior probability value indicating a shared variant.

Data visualization
A Manhattan plot was created with the GWASinspector package (version 1.5.1)[67]. Gene-based analysis (as computed by MAGMA) and gene Manhattan plot were performed and visualized with FUMA (v1.3.6)

SNP-to-Gene function[70,96]. Forest plots of the odds ratios and confidence intervals of all meta-GWAS hits were constructed in R (version 3.6.1) using the rmeta package (https://CRAN.R-project.org/package = rmeta, version 3.0). Regional association plots of each SNP were created with Locuszoom (version 0.13.2)[97].

Reporting summary
Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability
The summary statistics data generated in this study have been deposited in the GWAS catalog under accession code GCST90301252. The raw genotype data are protected and are not available due to data privacy laws. The open source data used in this study from FinnGen (v6) [https://r6.finngen.fi/pheno/M13_DUPUTRYEN], 1000Genomes (phase 1 and 3) [https://www.internationalgenome.org/data/], and the Haplotype Reference Consortium [http://www.haplotype-reference-consortium.org/] were downloaded from each of their respective websites. SNPs of interest were extracted with VCFtools [https://vcftools.github.io/]. GWAS Catalog database (version 21 April 2021) was used for in silico pleiotropy analysis. Single cell RNA sequencing data were aqcuired via correspondence. Blood cis-eQTL data were downloaded from the eQTLGen consortium [https://eqtlgen.org/]. Fibroblast cis-eQTL data were downloaded from the Genotype-Tissue Expression (GTEx) version 8 [https://gtexportal.org/home/]. Co-regulation analysis was performed using DEPICT and its accompanying expression dataset of 77,840 samples [https://github.com/perslab/depict]. The Phenoscanner database (version 2) was queried to look up quantitative trait loci (QTL) associations [http://www.phenoscanner.medschl.cam.ac.uk/login/?next = /data/]. GeneMANIA was used to construct composite networks of the prioritized genes based on the database accompanied by the software (build 12-02-2019) [http://genemania.org/]. The STRING database v11.0 to find the protein-protein interactions [https://version-11-0.string-db.org/]. The Genotype-Tissue Expression (GTEx) database (v8) was used to study gene expression of prioritized genes [https://gtexportal.org/home/]. To assess enrichment of tissue-specific genes, the Human Protein Atlas (PMID 25613900) and mouse gene expression as well as RNAseq data from the GTEx database were used [https://www.proteinatlas.org/].

References
1. Wynn, T. A. & Ramalingam, T. R. Mechanisms of fibrosis: therapeutic translation for fibrotic disease. *Nat. Med.* **18**, 1028–1040 (2012).
2. Kuo, R. Y. L., Ng, M., Prieto-Alhambra, D. & Furniss, D. Dupuytren's disease predicts increased all-cause and cancer-specific mortality: analysis of a large cohort from the U.K. clinical practice research datalink. *Plast. Reconstr. Surg.* **145**, 574e–582e (2020).
3. Lanting, R., Broekstra, D. C., Werker, P. M. N. & van den Heuvel, E. R. A systematic review and meta-analysis on the prevalence of Dupuytren disease in the general population of western countries. *Plast. Reconstr. Surg.* **133**, 593–603 (2014).
4. Wilburn, J., McKenna, S. P., Perry-Hinsley, D. & Bayat, A. The impact of Dupuytren disease on patient activity and quality of life. *J. Hand Surg. Am.* **38**, 1209–1214 (2013).
5. van Rijssen, A. L., ter Linden, H. & Werker, P. M. N. Five-year results of a randomized clinical trial on treatment in Dupuytren's disease: percutaneous needle fasciotomy versus limited fasciectomy. *Plast. Reconstr. Surg.* **129**, 469–477 (2012).
6. Layton, T. & Nanchahal, J. Recent advances in the understanding of Dupuytren's disease [version 1; referees: 3 approved]. *F1000 Res.* **8**, 1–8 (2019).
7. Alser, O. H., Kuo, R. Y. L. & Furniss, D. Nongenetic factors associated with Dupuytren's disease: a systematic review. *Plast. Reconstr. Surg.* **74**, 799–807 (2020).
8. Major, M. et al. Integrative analysis of Dupuytren's disease identifies novel risk locus and reveals a shared genetic etiology with BMI. *Genet. Epidemiol.* **43**, 629–645 (2019).
9. Majeed, M., Wiberg, A., Ng, M., Holmes, M. V. & Furniss, D. The relationship between body mass index and the risk of development of Dupuytren's disease: a mendelian randomization study. *J. Hand Surg. Eur. Vol.* **46**, 406–410 (2021).
10. Burkard, T. et al. The association of bariatric surgery and Dupuytren's disease: a propensity score-matched cohort study. *J. Hand Surg. Eur. Vol.* **47**, 288–295 (2022).
11. Dolmans, G. H. et al. Wnt signaling and Dupuytren's disease. *N. Engl. J. Med.* **365**, 307–317 (2011).
12. Ng, M. et al. A genome-wide association study of Dupuytren disease reveals 17 additional variants implicated in fibrosis. *Am. J. Hum. Genet.* **101**, 417–427 (2017).
13. Ågren, R. et al. Major genetic risk factors for Dupuytren's disease are inherited from Neandertals. *Mol. Biol. Evol.* **40**, 1–11 (2023).
14. Larsen, S. et al. Genetic and environmental influences in Dupuytren's disease. *J. Hand Surg. Eur. Vol.* **40**, 171–176 (2015).
15. The FinnGen, Project. *FinnGen Release 6* https://r6.finngen.fi/pheno/M13_DUPUTRYEN (2022).
16. Rentzsch, P., Witten, D., Cooper, G. M., Shendure, J. & Kircher, M. CADD: Predicting the deleteriousness of variants throughout the human genome. *Nuc. Acids Res.* **47**, D886–D894 (2019).
17. Sato, M. Upregulation of the Wnt/β-catenin pathway induced by transforming growth factor-β in hypertrophic scars and keloids. *Acta. Derm. Venereol.* **86**, 300–307 (2006).
18. Zhu, Z. et al. Shared genetic and experimental links between obesity-related traits and asthma subtypes in UK Biobank. *J. Allergy Clin. Immunol.* **145**, 537–549 (2020). Feb.
19. Zhu, Z. et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat. Genet.* **48**, 481–487 (2016).
20. Green, H., et al. A genome-wide association study identifies 5 loci associated with frozen shoulder and implicates diabetes as a causal risk factor. *PLoS Genet.* **10**, 100–132 (2021).
21. Yengo, L. et al. Meta-analysis of genome-wide association studies for height and body mass index in ~700 000 individuals of European ancestry. *Hum. Mol. Genet.* **27**, 3641–3649 (2018).
22. Lek, M. et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291 (2016).
23. Layton, T. B. et al. A vasculature niche orchestrates stromal cell phenotype through PDGF signaling: Importance in human fibrotic disease. *Proc. Natl Acad. Sci. USA* **119**, 1–11 (2022).
24. Ten Dam, E. J. P. M., van Beuge, M. M., Bank, R. A. & Werker, P. M. N. Further evidence of the involvement of the Wnt signaling pathway in Dupuytren's disease. *J. Cell Commun. Sign.* **10**, 33–40 (2016).
25. Piersma, B. et al. YAP1 is a driver of myofibroblast differentiation in normal and diseased fibroblasts. *Am. J. Pathol.* **185**, 3326–3337 (2015).
26. Piersma, B., Bank, R. A. & Boersema M. Signaling in fibrosis: TGF-β, WNT, and YAP/TAZ converge. *Front. Med.* **88**, e00521 (2021).
27. Hu, L., Lin, X., Lu, H., Chen, B. & Bai, Y. An overview of hedgehog signaling in fibrosis. *Mol. Pharm.* **87**, 174–182 (2015).
28. Omenetti, A. et al. Hedgehog signaling regulates epithelial-mesenchymal transition during biliary fibrosis in rodents and humans. *J. Clin. Investig.* **118**, 3331–3342 (2008).
29. Fitch, P. M., Howie, S. E. M. & Wallace, W. A. H. Oxidative damage and TGF-β differentially induce lung epithelial cell sonic hedgehog and tenascin-C expression: Implications for the regulation of lung remodelling in idiopathic interstitial lung disease. *Int J. Exp. Pathol.* **92**, 8–17 (2011).
30. Shen, X., Peng, Y. & Li, H. The injury-related activation of hedgehog signaling pathway modulates the repair-associated inflammation in liver fibrosis. *Front. Immunol.* **8**, 1450 (2017).

31. Dennler, S. et al. Induction of sonic hedgehog mediators by transforming growth factor-beta: smad3-dependent activation of Gli2 and Gli1 expression in vitro and in vivo. *Cancer Res.* **67**, 6981–6986 (2007).

32. Berndt, A., Kosmehl, H., Katenkamp, D. & Tauchmann, V. Appearance of the myofibroblastic phenotype in Dupuytren's disease is associated with a fibronectin, laminin, collagen type IV and tenascin extracellular matrix. *Pathobiology* **62**, 55–58 (1994).

33. Fu, H. et al. Tenascin-C Is a major component of the fibrogenic niche in kidney fibrosis. *J. Am. Soc. Nephrol.* **28**, 785–801 (2017).

34. Che, J. et al. Decreased expression of Dlg5 is associated with a poor prognosis and epithelial-mesenchymal transition in squamous cell lung cancer. *J. Thorac. Dis.* **13**, 3115–3125 (2021).

35. Chong, Y. C., Mann, R. K., Zhao, C., Kato, M. & Beachy, P. A. Bifurcating action of smoothened in hedgehog signaling is mediated by Dlg5. *Genes Dev.* **29**, 262–276 (2015).

36. Xie, J. Hedgehog signaling in prostate cancer. *Future Oncol.* **1**, 331–338 (2005).

37. Hidalgo, M. & Maitra, A. The Hedgehog pathway and pancreatic cancer. *N. Engl. J. Med.* **361**, 2094–2096 (2009).

38. Xie, J. Molecular biology of basal and squamous cell carcinomas. *Adv. Exp. Med. Biol.* **624**, 241–251 (2008).

39. Bray, S. J. Notch signalling in context. *Nat. Rev. Mol. Cell Biol.* **17**, 722–735 (2016).

40. Fan, J. et al. Targeting the Notch and TGF-β signaling pathways to prevent retinal fibrosis in vitro and in vivo. *Theranostics* **10**, 7956–7973 (2020).

41. Zhang, K. et al. The liver-enriched lnc-LFAR1 promotes liver fibrosis by activating TGFβ and Notch pathways. *Nat. Commun.* **8**, 144 (2017).

42. Edeling, M., Ragi, G., Huang, S., Pavenstädt, H. & Susztak, K. Developmental signalling pathways in renal fibrosis: the roles of Notch, Wnt and Hedgehog. *Nat. Rev. Nephrol.* **12**, 426–439 (2016).

43. Bielesz, B. et al. Epithelial Notch signaling regulates interstitial fibrosis development in the kidneys of mice and humans. *J. Clin. Investig.* **120**, 4040–4054 (2010).

44. Hicks, C. et al. Fringe differentially modulates Jagged1 and Delta1 signalling through Notch1 and Notch2. *Nat. Cell Biol.* **2**, 515–520 (2000).

45. Tian, J. et al. Loss of CHSY1, a secreted FRINGE enzyme, causes syndromic brachydactyly in humans via increased NOTCH signaling. *Am. J. Hum. Genet. Inter.* **87**, 768–778 (2010).

46. Zhao, B. et al. TEAD mediates YAP-dependent gene induction and growth control. *Genes Dev.* **22**, 1962–1971 (2008).

47. Smith, S. P., Devaraj, V. S. & Bunker, T. D. The association between frozen shoulder and Dupuytren's disease. *J. Should. Elb. Surg.* **10**, 149–151 (2001).

48. Itoh, Y. et al. A common SNP risk variant MT1-MMP causative for Dupuytren's disease has a specific defect in collagenolytic activity.pdf. *Matrix Biol.* **97**, 20–39 (2021).

49. Hutchinson, J. W., Tierney, G. M., Parsons, S. L. & Davis, T. R. Dupuytren's disease and frozen shoulder induced by treatment with a matrix metalloproteinase inhibitor. *J. Bone Jt. Surg. Br.* **80**, 907–908 (1998).

50. van Beuge, M. M., ten Dam, E. J. P. M., Werker, P. M. N. & Bank, R. A. Wnt pathway in Dupuytren disease: connecting profibrotic signals. *Transl. Res.* **166**, 762–771.e3 (2015).

51. Riesmeijer, S. A. et al. Polygenic risk associations with clinical characteristics and recurrence of Dupuytren's disease. *Plast Reconstr Surg.* (2023). Online ahead of print.

52. Becker, K. et al. The importance of genetic susceptibility in Dupuytren's disease. *Clin. Genet.* **87**, 483–487 (2015).

53. Scholtens, S. et al. Cohort profile: lifeLines, a three-generation cohort study and biobank. *Int J. Epidemiol.* **44**, 1172–1180 (2015).

54. Becker, K. et al. Meta-analysis of genome-wide association studies and network analysis-based integration with gene expression data identify new suggestive loci and unravel a Wnt-centric network associated with Dupuytren's disease. *PLoS One* **11**, 1–18 (2016).

55. Bycroft, C. et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).

56. Holle, R., Happich, M., Löwel, H. & Wichmann, H. E. KORA–a research platform for population based health research. *Gesundheitswesen* **67**, S19–S25 (2005).

57. Nöthlings, U. & Krawczak, M. PopGen. A population-based biobank with prospective follow-up of a control group. *Bundesgesundheitsbl. Gesundheitsforsch. Gesundheitssch.* **55**, 831–835 (2012).

58. Dolmans, G. H., de Bock, G. H. & Werker, P. M. Dupuytren diathesis and genetic risk. *J. Hand Surg.* **37**, 2106–2111 (2012).

59. Laura, Fumagalli. & Nick, Buck. *Understanding Society* 14th edn, Vol. 6 (University of Essex Institute for Social and Economic Research, 2017).

60. Layton, T. B. et al. Cellular census of human fibrosis defines functionally distinct stromal cell types and states. *Nat. Commun.* **11**, 1–11 (2020).

61. Purcell, S. et al. PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).

62. Auton, A. et al. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).

63. Zheng, X. & Davis, J. W. SAIGEgds - an efficient statistical tool for large-scale PheWAS with mixed models. *Bioinformatics* **37**, 728–730 (2021).

64. McCarthy, S. et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* **48**, 1279–1283 (2016).

65. Purcell, S. et al. PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).

66. Marchini, J., Howie, B., Myers, S., McVean, G. & Donnelly, P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat. Genet.* **39**, 906–913 (2007).

67. Ani, A., van der Most, P., Snieder, H., Vaez, A. & Nolte, I. GWASinspector: comprehensive quality control of genome-wide association study results. *Bioinformatics* **37**, 129–130 (2020).

68. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).

69. Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55**, 997–1004 (1999).

70. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* **8**, 1–10 (2017).

71. Vaez, A. et al. In silico post genome-wide association studies analysis of C-reactive protein Loci suggests an important role for interferons. *Circ. Cardiovasc. Genet.* **8**, 487–497 (2015).

72. Danecek, P. et al. The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).

73. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nuc. Acids Res.* **38**, e164 (2010).

74. Buniello, A. et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nuc. Acids Res.* **47**, D1005–D1012 (2019).

75. Kamali, Z., Keaton, J. M. & Javanmard, S. H. Large-scale multi-omics studies provide new insights into blood pressure regulation. *Int J. Mol. Sci.* **23**, 7557 (2022).

76. Asefa, N. G. et al. Bioinformatic prioritization and functional annotation of GWAS-based candidate genes for primary open-angle Glaucoma. *Genes Basel* **13**, 1055 (2022).

77. Benner, C. et al. FINEMAP: efficient variable selection using summary data from genome-wide association studies. *Bioinformatics* **32**, 1493–1501 (2016).

78. Qi, T. et al. Identifying gene targets for brain-related traits using transcriptomic and methylomic data from blood. *Nat. Commun.* **9**, 2282 (2018).

79. Slowikowski, K., Hu, X. & Raychaudhuri, S. SNPsea: an algorithm to identify cell types, tissues and pathways affected by risk loci. *Bioinformatics* **30**, 2496–2497 (2014).

80. Galili, T., O'Callaghan, A., Sidi, J. & Sievert, C. heatmaply: an R package for creating interactive cluster heatmaps for online publishing. *Bioinformatics* **34**, 1600–1602 (2018).

81. Zheng Z. et al. Leveraging functional genomic annotations and genome coverage to improve polygenic prediction of complex traits within and between ancestries. *bioRxiv* https://www.biorxiv.org/10.1101/2022.01.07.475305v1 (2022).

82. Lee, S. H., Goddard, M. E., Wray, N. R. & Visscher, P. M. A better coefficient of determination for genetic profile analysis. *Genet Epidemiol.* **36**, 214–224 (2012).

83. Riesmeijer, S. A., Werker, P. M. N. & Nolte, I. M. Ethnic differences in prevalence of Dupuytren disease can partly be explained by known genetic risk variants. *Eur. J. Hum. Genet.* **27**, 1876–1884 (2019).

84. Willer, C. et al. Discovery and refinement of loci associated with lipid levels. *Nat. Genet* **45**, 1274–1283 (2013).

85. Mahajan, A. et al. Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *Nat. Genet.* **50**, 1505–1513 (2018).

86. Wellcome Sanger. *The Haplotype Reference Consortium* http://www.haplotype-reference-consortium.org/ (2022).

87. Chen, J. et al. The trans-ancestral genomic architecture of glycemic traits. *Nat. Genet.* **53**, 840–860 (2021).

88. Allen, R. J. et al. Genome-wide association study of susceptibility to idiopathic pulmonary fibrosis. *Am. J. Respir. Crit. Care Med.* **201**, 564–574 (2020).

89. López-Isac, E. et al. GWAS for systemic sclerosis identifies multiple risk loci and highlights fibrotic and vasculopathy pathways. *Nat. Commun.* **10**, 4955 (2019).

90. Hindorff, L. A. et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl Acad. Sci. USA* **106**, 9362–9367 (2009).

91. Zheng, I. et al. LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics* **33**, 272–279 (2017).

92. Morris, J. A. et al. An atlas of genetic influences on osteoporosis in humans and mice. *Nat. Genet.* **51**, 258–266 (2019).

93. Yengo, L. et al. A saturated map of common genetic variants associated with human height from 5.4 million individuals of diverse ancestries. *bioRxiv* https://www.biorxiv.org/10.1101/2022.01.07.475305v1 (2022).

94. Bulik-Sullivan, B. et al. An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241 (2015).

95. Wallace, C. A more accurate method for colocalisation analysis allowing for multiple causal variants. *PLoS Genet.* **17**, e1009440 (2021).

96. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput. Biol.* **11**, 1–19 (2015).

97. Pruim, R. J. et al. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* **26**, 2336–2337 (2010).

Author contributions

I.N. conceived the study. S.R., I.N., Z.K., and A.V. were involved in protocol development. T.L. and J.N. provided data required to perform part of this work. S.R., I.N., Z.K., and A.V. were involved in data analysis. SR, IN, and DF wrote the first draft of the manuscript. All authors reviewed and edited the manuscript and approved the final version of the manuscript. For the Dutch cohorts: S.R., I.N., B.P. and P.W. participated in study concept and design. S.R. and P.W. participated in phenotype data acquisition and/or quality control. S.R. and I.N. participated in genotype data acquisition and/or quality control. S.R. and I.N. participated in data analysis and interpretation. For the United Kingdom cohorts: M.Ng. and D.F. participated in study concept and design. M.Ng. and D.F. participated in phenotype data acquisition and/or quality control. M.Ng. and D.F. participated in genotype data acquisition and/or quality control. M.Ng. and D.F. participated in data analysis and interpretation. For the German cohorts: H.H. and K.B. participated in study concept and design. H.H. and K.B. participated in phenotype data acquisition and/or quality control. H.H., M.No., D.D., and K.B. participated in genotype data acquisition and/or quality control. H.H., M.No., D.D., and K.B. participated in data analysis and interpretation.

Competing interests

The authors have the following to disclose: P.W. was science and engineering research board (SERB) member for an industry sponsored study on collagenase of Fidia Ltd, Milan, Italy. J.N. consults for and holds equity in 180 Life Sciences, which has exclusively licensed IP from the University of Oxford on the use of anti-TNF for early-stage Dupuytren's disease. All other authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-023-44451-0.

Correspondence and requests for materials should be addressed to Sophie A. Riesmeijer.

Peer review information *Nature Communications* thanks David Karasik and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.