A Genome-wide Association Study of Dupuytren Disease Reveals 17 Additional Variants Implicated in Fibrosis

Michael Ng,1 Dipti Thakkar,1 Lorraine Southam,2,3 Paul Werker,4 Roel Ophoff,5 Kerstin Becker,6,7 Michael Nothnagel,9 Andre Franke,8 Peter Nürnberg,6,7 Ana Isabel Espirito-Santo,1 David Izadi,1 Hans Christian Hennies,6,7,9 Jagdeep Nanchahal,1,10 Eleftheria Zeggini,2 and Dominic Furniss1,10,11,*

Individuals with Dupuytren disease (DD) are commonly seen by physicians and surgeons across multiple specialties. It is an increasingly common and disabling fibroproliferative disorder of the palmar fascia and the most common inherited disorder of the connective tissue. It is the most frequent example of a tissue-specific fibrotic disease: others include pulmonary, renal, hepatic, and skin fibrosis. It is accepted that there are common features of all fibrotic diseases, but some pathologic pathways are likely to be tissue specific.1

DD is characterized by the initial development of myofibroblast-rich nodules in the palm of the hand. These myofibroblasts express alpha-smooth muscle actin (α-SMA) and secrete types III and I collagen, leading to the formation of abnormal cords in the palm of the hand. In a proportion of people with DD, the myofibroblasts cause contraction, leading to flexion contractures of the involved digits and subsequent functional impairment.2,3 As the hand is the sensorimotor end-organ of the upper limb, impairment here has a disproportionate effect on the quality of life of the individual.4 Additionally, because DD is associated with other forms of fibrosis, it may serve as an ideal human model system for fibrotic disease, and the routine excision of tissue as a part of treatment facilitates experimental medicine studies.5

Introduction

Dupuytren disease (DD [MIM: 126900]) is a progressive fibroproliferative disease of the palmar fascia and the most common inherited disorder of the connective tissue. It is the most frequent example of a tissue-specific fibrotic disease: others include pulmonary, renal, hepatic, and skin fibrosis. It is accepted that there are common features of all fibrotic diseases, but some pathologic pathways are likely to be tissue specific.1

DD is very common, affecting 5%–25% of people in populations of European descent, and there is evidence that the prevalence is increasing.6–8 The mainstay of treatment for DD is surgery, though newer modalities are increasing in popularity. Despite this, complications and recurrence of disease are both common, even after adequate primary treatment.9,10

DD has a substantial heritable component. A twin study from Denmark estimated the heritability of DD at 80%11 and a sibling recurrence study from the UK estimated the λ_S to be 4.48, confirming a strong genetic predisposition to DD.12 Furthermore, age at first surgical intervention is significantly younger in those with a positive family history.13 Similarly, there is evidence that multiple non-genetic factors, such as smoking, alcohol intake, diabetes, and hyperlipidemia, also play a role in disease development.14

We have previously undertaken a pilot GWAS in 960 Dutch DD-affected individuals to begin to delineate the common genetic variation underlying this predisposition. This defined nine susceptibility loci and revealed the hitherto unsuspected importance of components of the WNT signaling pathway in the pathogenesis of DD.15 To boost power for the detection of common-frequency signals, here we undertook a 4-fold larger GWAS in 3,871 UK

1Nuffield Department of Orthopaedics, Rheumatology, and Musculoskeletal Science, University of Oxford, Botnar Research Centre, Windmill Road, Oxford OX3 7HE, UK; 2Wellcome Trust Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge CB10 1SA, UK; 3Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, UK; 4University of Groningen, University Medical Centre Groningen, Department of Plastic Surgery, Hanzeplein 1, 9713 GZ Groningen, the Netherlands; 5UCLA Center for Neurobehavioral Genetics, 695 Charles E. Young Drive South, Los Angeles, CA 90095, USA; 6Cologne Center for Genomics, University of Cologne, Weyertal 115b, 50931 Köln, Germany; 7Cluster of Excellence on Cellular Stress Responses in Aging-associated Diseases, University of Cologne, 50931 Köln, Germany; 8Institute of Clinical Molecular Biology, Christian-Albrechts-University of Kiel, University Hospital Schleswig-Holstein, 24105 Kiel, Germany; 9Department of Biological Sciences, University of Huddersfield, Huddersfield HD1 3DH, UK; 10Department of Plastic and Reconstructive Surgery, Oxford University Hospitals NHS Foundation Trust, John Radcliffe Hospital, Oxford OX3 9DU, UK; 11NIHR Biomedical Research Centre, NDORMS, University of Oxford, Botnar Research Centre, Windmill Road, Oxford OX3 7HE, UK
*Correspondence: dominic.furniss@ndorms.ox.ac.uk
http://dx.doi.org/10.1016/j.ajhg.2017.08.006
© 2017 The Author(s). This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).
individuals with surgically validated DD. Replication of significant and suggestive loci was performed in a total of 4,041 surgically validated DD-affected case subjects from the UK, the Netherlands, and Germany.

Material and Methods

Ethical Approval
The study was approved by the Research Ethics Committee or equivalent at all institutions where the work was carried out: Oxfordshire Research Ethics Committee B/09/H0605/65 for the British Society for Surgery of the Hand Genetics of Dupuytren’s Disease (BSSH-GODD) study (UK), Medical Ethics Committee (METc) 2007/067 for the Genetic Origin of Dupuytren Disease (GODDAF) Study (the Netherlands), and University of Cologne 14/292 for the German Dupuytren Study (Germany). Informed consent was obtained from all subjects.

Phenotype Definition and Study Populations
We used samples from three European countries for this study. In all cohorts, the DD-affected case subjects were individuals who had undergone surgical treatment for their disease. The UK cohort consisted of a total of 5,408 case subjects from the BSSH-GODD Study and 9,961 population-based control subjects from the United Kingdom Household Longitudinal Study (UKHLS), which were divided into 4,891 control subjects for the discovery phase and 5,070 control subjects for the replication phase. The Dutch cohort consisted of 2,195 case subjects from the GODDAF Study and 1,983 control subjects from the Lifelines cohort study. The German cohort consisted of 768 case subjects from the German Dupuytren Study and 1,353 control subjects from the PopGen and KORA studies. The cohorts included all samples analyzed in our previous GWAS.15

Biological Samples
For the BSSH-GODD cohort, salivary samples were collected using the Oragene-OG250 salivary DNA collection kit (DNA Genotek). DNA was extracted according to manufacturer’s instructions and stored at −80°C. Diseased fascial samples removed at surgery were immediately placed in EMEM media (Lonza) and transferred by overnight courier to our laboratory.

The UK Household Longitudinal Study is a stratified clustered random sample of households representative of the UK population, led by the Institute for Social and Economic Research at the University of Essex and funded by the Economic and Social Research Council. Blood was taken and DNA isolated by standard methods. The genome-wide scan data were analyzed and deposited by the Wellcome Trust Sanger Institute. Information on how to access the data can be found on the Understanding Society website.

For the GODDAF study, case subjects were identified from plastic surgery clinics within the Netherlands, and DNA and phenotype data were obtained as previously described.15

LifeLines is a population-based cohort study based in the Netherlands and has been previously described.16 For the purpose of this study, DNA samples from participants, age- and sex-matched to the GODDAF case subjects, were isolated (project number OV14_0257).

For the German Dupuytren study cohort, blood samples were collected from case subjects from Germany and Switzerland by the German Dupuytren Study Group13 and DNA was extracted with standard procedures. 1,353 control subjects were obtained from the Popgen and KORA studies.

Genotyping, Association Analysis, and Imputation
We genotyped 4,201 UK DD-affected case subjects using Illumina HumanCoreExome arrays at the Wellcome Trust Sanger Institute comprising 538,448 SNPs. The data were called using the Illumina GenCall algorithm. Quality control and association analyses were performed in PLINK v.1.9 and R v.3.3.1. We initially performed sample-level quality control (Figure S1). Briefly, we first removed all SNPs with a call rate < 90%. We standardized the output data to NCBI build 37 (hg19) and the strand alignment using scripts provided by Dr. William Rayner. We then removed 242 samples with one or more of the following properties: call rate < 98%; heterozygosity > 3 standard deviations from the mean; different genotype-derived sex and reported sex; or failure of genotyped SNPs to match the pre-GWAS Sequenom fingerprinting. We merged our data with publicly available data from the 1000 Genomes Project and performed principal components analysis (PCA) to define (and remove from further analysis) those people who were ethnic outliers by visual inspection (Figure S2).

We then performed SNP-level quality control on this sample set. Briefly, we excluded SNPs with call rate < 98%, those with Hardy-Weinberg equilibrium (HWE) p < 0.0001, and those with a cluster separation score of < 0.4. We also removed non-autosomal SNPs and those that were duplicated.

This generated a final set of 3,959 case subjects genotyped at 494,982 SNPs. From the UKHLS control subjects we selected 4,891 individuals genotyped at 525,314 SNPs after identical quality control. We then combined these control subjects with our case subjects. From this combined dataset, we further excluded 86 case subjects and 202 control subjects from a total of 604 related individuals by average identity-by-descent allele sharing (PiHAT ≥ 0.185 in PLINK), 1 sample due to poor genotype calling, and a further 4 ethnic outliers, leaving 3,871 case subjects and 4,686 control subjects for the association analysis. Here we report on the analysis of common variants within this cohort: 238,825 SNPs with minor allele frequency ≥ 0.05, as less common variants were poorly called.

For the discovery phase, we performed association analysis using logistic regression with sex and the first two principal component (PC1 and PC2) of the PCA as covariates. We did not use further principal components in our regression model as after adjustment for PC2, we saw no further separation of distinct subsets, and the genomic inflation factor did not decrease further. We then performed SNP-level quality control on this sample set, removing SNPs with call rate < 98%, those with Hardy-Weinberg equilibrium (HWE) p < 0.0001, and those with a cluster separation score of < 0.4. We also removed non-autosomal SNPs and those that were duplicated.

For the discovery phase, we performed association analysis using logistic regression with sex and the first two principal component (PC1 and PC2) of the PCA as covariates. We did not use further principal components in our regression model as after adjustment for PC2, we saw no further separation of distinct subsets, and the genomic inflation factor did not decrease further. We then performed SNP-level quality control on this sample set, removing SNPs with call rate < 98%, those with Hardy-Weinberg equilibrium (HWE) p < 0.0001, and those with a cluster separation score of < 0.4. We also removed non-autosomal SNPs and those that were duplicated.

For the discovery phase, we performed association analysis using logistic regression with sex and the first two principal component (PC1 and PC2) of the PCA as covariates. We did not use further principal components in our regression model as after adjustment for PC2, we saw no further separation of distinct subsets, and the genomic inflation factor did not decrease further. We then performed SNP-level quality control on this sample set, removing SNPs with call rate < 98%, those with Hardy-Weinberg equilibrium (HWE) p < 0.0001, and those with a cluster separation score of < 0.4. We also removed non-autosomal SNPs and those that were duplicated.

For the discovery phase, we performed association analysis using logistic regression with sex and the first two principal component (PC1 and PC2) of the PCA as covariates. We did not use further principal components in our regression model as after adjustment for PC2, we saw no further separation of distinct subsets, and the genomic inflation factor did not decrease further. We then performed SNP-level quality control on this sample set, removing SNPs with call rate < 98%, those with Hardy-Weinberg equilibrium (HWE) p < 0.0001, and those with a cluster separation score of < 0.4. We also removed non-autosomal SNPs and those that were duplicated.
Sequenom MassARRAY platform. German case and control subjects were previously genotyped on the Affymetrix Human SNP Array 6.0. Where no direct or tag SNP was available on the Affymetrix platform, German case and control subjects were genotyped using TaqMan probes (Table S1). SNPs with call rate < 90% or SNPs with deviation from HWE (p < 0.0001) were removed, leaving 46 (31 in the German cohort, 41 in the Netherlands cohort, and 42 in the UK cohort) in the final dataset. 246 samples were removed due to call rate < 90%, and 70 samples were removed due to sex mismatch between self-reported data and genotyping result. SNP rs2598107 was separately replicated only on the UK cohort, 38 samples of which were removed due to call rate < 90%. The remaining UKHLS control subjects were again genotyped on the Illumina HumanCoreExome platform and underwent QC as described above. We used multiple genotyping platforms in the replication phase, so constructed Forest plots in R to check for heterogeneity. Since our replication signals were in the same direction and of similar magnitude to our discovery results, it is unlikely that genotyping artifact was responsible for the observed associations (Figure S5).

For association analysis of the replication phase, we performed logistic regression and used sex as a covariate. The Breslow-Day test was used to test for heterogeneity. We performed a fixed-effects meta-analysis of discovery and replication phase using the inverse variance method, assuming all studies share a common true effect size at each locus. The explained heritability for regions that contained two index SNPs, we identified the association region until the total reached 0.99.

For imputation, we phased our dataset using SHAPEIT2. The phased dataset was submitted to the Haploype Reference Consortium imputation service, utilizing the Sanger server and the standard PBWT pipeline. Imputed data were subjected to quality control. We removed SNPs with info score < 0.3, MAF < 1%, or significant deviation from Hardy Weinberg equilibrium (p < 1 × 10^-6). We used SNPTEST v2.5.2 to calculate the Bayes factor for each SNP with the assumption that there was only one causal SNP per associated region, and the additive model of prior distribution. For regions that contained two index SNPs, we identified the BF for the second index SNP by conditioning on the first index SNPs. Posterior probability was defined as Bayes factor for SNP k divided by the summation of the BF for every SNP in the selected region, 500 kb upstream and downstream of the index SNP, as previously described. 99% credible sets were constructed by summing the ranked posterior probability of every SNP within each associated region until the total reached 0.99.

Tissue Culture

Primary cells were disaggregated from fresh surgical tissue samples using 300 units/g type II collagenase at 1 mg/mL (Worthington Chemical) in DMEM (Lonza) supplemented with 5% FBS (Labtech) overnight at 37°C with 5% CO2. After incubation, cells were filtered using 40 μm culture plate meshes, pelleted, and cultured on 10 cm² Petri dishes. Primary myofibroblasts were cultured in DMEM supplemented with 10% FBS (Labtech), 1% penicillin/streptomycin, and 1× Glutamax (ThermoFisher Scientific).

Immunocytochemistry

Diseased, surgically resected palmar fascia was disaggregated as previously described, and myofibroblasts were seeded on 35 mm FluoroDish tissue culture dishes (World Precision Instruments) at 50,000 cells per dish. Cells were fixed in 4% formalde-
In the GWAS discovery phase, we used the Illumina HumanCoreExome array to test 238,825 common-frequency variants (MAF ≥ 0.05) for association with DD in 3,871 UK case subjects and 4,686 UK control subjects, after quality control. This yielded genome-wide significant associations (p ≤ 5 × 10⁻⁸) at 14 variants, including 8 of the 9 previously reported loci (Figure 1).

In the replication phase, we genotyped 45 SNPs with p ≤ 1 × 10⁻⁵ in the discovery set, in a total of 4,041 case subjects and 8,251 control subjects from the UK, the Netherlands, and Germany. In addition, there were a further four SNPs with suggestive association (p ≤ 1 × 10⁻⁵) and one with genome-wide significant association (rs246105, frequency 20.1%, OR = 0.796, p = 1.34 × 10⁻⁸) for which we were unable to design an appropriate assay (Tables 1 and S1). After fixed-effects meta-analysis, we confirmed association at all 9 previously reported loci and defined 15 further loci with genome-wide significant evidence of association (Table 1; Figure S6).

Conditional analysis at all associated loci confirmed two independent signals at two loci. On chromosome 7, after conditioning on rs16879765, rs2598107 showed residual evidence of association (r² = 0.0, frequency 44.7%, OR = 1.48, pcond = 6.85 × 10⁻¹¹). Similarly, on chromosome 8, after conditioning on rs629535, rs2912522 showed residual evidence of association (r² = 0.0, frequency 19.7%, OR = 0.73, pcond = 1.31 × 10⁻¹⁴; Table 1).

To further characterize the genetic architecture of DD, we tested first the contribution of all autosomal common-frequency variants (MAF ≥ 0.05) and second the 26 genome-wide significant variants alone to trait variance using genome-wide complex trait analysis (GCTA) and estimated them to be 53.1% and 11.3%, respectively.

Imputation and Construction of 99% Credible Sets
We imputed our dataset using the Haplotype Reference Consortium resource. We calculated single SNP Bayes Factors (BF) for 7,218,238 SNPs within our imputed dataset, containing variants that passed our QC criteria. Variants with the highest BF within each associated region from our meta-analysis were used as the index SNP for the construction of 99% credible sets. Similar to our primary
Table 1. SNPs Significantly Associated with Dupuytren Disease (p $\leq 5 \times 10^{-8}$)

<table>
<thead>
<tr>
<th>Chromosome</th>
<th>Position</th>
<th>rsID</th>
<th>Allele</th>
<th>EAFa</th>
<th>Discovery</th>
<th>Replication</th>
<th>Meta-analysis</th>
<th>Selected Nearby Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>22698447</td>
<td>rs7524102</td>
<td>G</td>
<td>0.214</td>
<td>7.68 x 10^-12</td>
<td>1.332</td>
<td>6.43 x 10^-5</td>
<td>1.448</td>
</tr>
<tr>
<td>1</td>
<td>162672011</td>
<td>rs17433710</td>
<td>C</td>
<td>0.12</td>
<td>9.13 x 10^-7</td>
<td>0.791</td>
<td>3.73 x 10^-5</td>
<td>0.833</td>
</tr>
<tr>
<td>5</td>
<td>108672946</td>
<td>rs246105c</td>
<td>T</td>
<td>0.201</td>
<td>1.34 x 10^-9</td>
<td>0.796</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>6</td>
<td>149797014</td>
<td>rs394563</td>
<td>G</td>
<td>0.411</td>
<td>2.03 x 10^-8</td>
<td>0.828</td>
<td>1.02 x 10^-10</td>
<td>0.827</td>
</tr>
<tr>
<td>7</td>
<td>3318658</td>
<td>rs10276303</td>
<td>T</td>
<td>0.26</td>
<td>2.89 x 10^-9</td>
<td>0.817</td>
<td>6.00 x 10^-9</td>
<td>0.831</td>
</tr>
<tr>
<td>7</td>
<td>37973014</td>
<td>rs2598107d</td>
<td>T</td>
<td>0.474</td>
<td>1.11 x 10^-10</td>
<td>1.478</td>
<td>1.82 x 10^-15</td>
<td>1.475</td>
</tr>
<tr>
<td>7</td>
<td>37989095</td>
<td>rs16879765</td>
<td>T</td>
<td>0.178</td>
<td>7.15 x 10^-11</td>
<td>1.926</td>
<td>2.82 x 10^-42</td>
<td>1.837</td>
</tr>
<tr>
<td>7</td>
<td>116892846</td>
<td>rs38904</td>
<td>C</td>
<td>0.464</td>
<td>1.02 x 10^-11</td>
<td>1.254</td>
<td>6.52 x 10^-13</td>
<td>1.253</td>
</tr>
<tr>
<td>8</td>
<td>25845674</td>
<td>rs10866846</td>
<td>A</td>
<td>0.421</td>
<td>3.14 x 10^-11</td>
<td>1.249</td>
<td>1.78 x 10^-6</td>
<td>1.148</td>
</tr>
<tr>
<td>8</td>
<td>69992380</td>
<td>rs2912522d</td>
<td>G</td>
<td>0.201</td>
<td>1.29 x 10^-16</td>
<td>0.72</td>
<td>3.26 x 10^-14</td>
<td>0.751</td>
</tr>
<tr>
<td>8</td>
<td>70007938</td>
<td>rs629535</td>
<td>T</td>
<td>0.351</td>
<td>2.84 x 10^-28</td>
<td>1.477</td>
<td>1.17 x 10^-15</td>
<td>1.275</td>
</tr>
<tr>
<td>8</td>
<td>109228008</td>
<td>rs617144</td>
<td>G</td>
<td>0.402</td>
<td>3.70 x 10^-19</td>
<td>0.737</td>
<td>9.92 x 10^-16</td>
<td>0.794</td>
</tr>
<tr>
<td>8</td>
<td>145504343</td>
<td>rs7838717</td>
<td>T</td>
<td>0.405</td>
<td>2.55 x 10^-6</td>
<td>1.173</td>
<td>3.91 x 10^-9</td>
<td>1.188</td>
</tr>
<tr>
<td>9</td>
<td>1201156</td>
<td>rs12342106</td>
<td>A</td>
<td>0.308</td>
<td>9.76 x 10^-12</td>
<td>1.289</td>
<td>6.40 x 10^-16</td>
<td>1.29</td>
</tr>
<tr>
<td>10</td>
<td>44842503</td>
<td>rs9525927</td>
<td>G</td>
<td>0.167</td>
<td>5.80 x 10^-6</td>
<td>0.823</td>
<td>6.76 x 10^-6</td>
<td>0.842</td>
</tr>
<tr>
<td>14</td>
<td>23312594</td>
<td>rs1042704</td>
<td>A</td>
<td>0.248</td>
<td>8.72 x 10^-13</td>
<td>1.326</td>
<td>1.12 x 10^-8</td>
<td>1.213</td>
</tr>
<tr>
<td>14</td>
<td>51074461</td>
<td>rs1032466</td>
<td>C</td>
<td>0.306</td>
<td>4.90 x 10^-9</td>
<td>0.812</td>
<td>6.82 x 10^-10</td>
<td>0.824</td>
</tr>
<tr>
<td>15</td>
<td>56229760</td>
<td>rs1509406</td>
<td>G</td>
<td>0.356</td>
<td>4.03 x 10^-6</td>
<td>1.175</td>
<td>1.41 x 10^-5</td>
<td>1.154</td>
</tr>
<tr>
<td>15</td>
<td>68628163</td>
<td>rs2306022</td>
<td>T</td>
<td>0.11</td>
<td>7.59 x 10^-6</td>
<td>1.286</td>
<td>2.57 x 10^-6</td>
<td>1.266</td>
</tr>
<tr>
<td>15</td>
<td>89238184</td>
<td>rs6496519</td>
<td>T</td>
<td>0.164</td>
<td>9.35 x 10^-8</td>
<td>0.795</td>
<td>1.42 x 10^-10</td>
<td>0.789</td>
</tr>
<tr>
<td>16</td>
<td>75506593</td>
<td>rs977987</td>
<td>G</td>
<td>0.403</td>
<td>6.24 x 10^-7</td>
<td>1.184</td>
<td>8.84 x 10^-5</td>
<td>1.12</td>
</tr>
<tr>
<td>18</td>
<td>9762933</td>
<td>rs9951109</td>
<td>C</td>
<td>0.133</td>
<td>1.24 x 10^-7</td>
<td>0.776</td>
<td>8.89 x 10^-5</td>
<td>0.852</td>
</tr>
<tr>
<td>19</td>
<td>57678194</td>
<td>rs11672517</td>
<td>A</td>
<td>0.284</td>
<td>1.42 x 10^-13</td>
<td>1.331</td>
<td>2.71 x 10^-5</td>
<td>1.384</td>
</tr>
<tr>
<td>20</td>
<td>38300807</td>
<td>rs6016142</td>
<td>T</td>
<td>0.132</td>
<td>1.19 x 10^-6</td>
<td>1.282</td>
<td>1.98 x 10^-8</td>
<td>1.274</td>
</tr>
<tr>
<td>20</td>
<td>39320751</td>
<td>rs6102095</td>
<td>A</td>
<td>0.125</td>
<td>8.54 x 10^-7</td>
<td>0.792</td>
<td>1.10 x 10^-15</td>
<td>0.71</td>
</tr>
<tr>
<td>22</td>
<td>46459132</td>
<td>rs7291412</td>
<td>T</td>
<td>0.413</td>
<td>1.24 x 10^-15</td>
<td>1.316</td>
<td>1.14 x 10^-16</td>
<td>1.269</td>
</tr>
</tbody>
</table>

aBased on human genome build hg19.

bThe effect allele frequency (EAF) in the total cohort is shown, except for rs246105, where the effect allele frequency in the discovery set is shown.

cWe were unable to design an assay for this SNP in the replication phase.

dIdentified by conditional analysis.

bThe effect allele frequency (EAF) in the total cohort is shown, except for rs246105, where the effect allele frequency in the discovery set is shown.

cWe were unable to design an assay for this SNP in the replication phase.

dIdentified by conditional analysis.
analysis, conditional analysis revealed the same two loci with two independent signals. On chromosome 7, after conditioning on rs117402009 (BF = 9.87 × 10^{45}), we found residual evidence of association for rs2598100 (BF_{cond} = 7.29 × 10^{-31}), and on chromosome 8, after conditioning on rs2472141 (BF = 2.22 × 10^{28}), we found residual association for rs2981040 (BF_{cond} = 1.75 × 10^{-11}). We therefore constructed independent credible sets based around each independent signal at these loci (Figure S7 and Table S2). Intriguingly, for one of the credible sets constructed, the genotyped SNP rs1042704 (BF = 1.67 × 10^{10}) had a posterior probability greater than 0.99 and therefore appears to be the causative allele at that locus. Overall, the credible sets range in size from 1 to 293 variants, with a median size of 27.5. Details of the 99% credible sets can be found in Table S2 and Figure S7.

rs16879765
We further investigated the functional consequences of the statistically most associated SNP from the direct genotyping, rs16879765, as a proof of principle that using myofibroblasts from surgically resected DD tissue could help define the causative gene at a particular locus. This SNP is located in an intron of the gene EPDR1 and approximately 4 kb upstream of SFRP4 (MIM: 606570) (Figure 2A). EPDR1 is a poorly characterized type II transmembrane protein that shares some homology with ependymins and protocadherins. EPDR1 has been shown to be upregulated in CD34^+ hematopoetic stem cells and colorectal cancer cells. SFRP4 is a secreted protein with homology to the membrane-bound WNT receptors FZD. It is thought to modulate WNT signaling by competing for WNT ligands with Frizzled receptors.

We utilized myofibroblasts up to passage three derived from surgically resected DD samples to study the genotype-specific expression of SFRP4 and EPDR1. The homozygous high-risk (TT) genotype at rs16879765 showed significantly greater SFRP4 expression compared to the low-risk (CC) genotype, with the heterozygous state showing intermediate expression levels. There was no genotype-specific differential expression of EPDR1 (Figure 2B). Immunocytochemistry performed on disaggregated primary cells from surgically resected DD tissue failed to demonstrate EPDR1 expression but showed cytoplasmic SFRP4 expression (Figure 2C). Furthermore, immunohistochemistry confirmed expression of SFRP4 in fixed surgically resected fibrotic DD tissue (Figure 2D), but not in palm skin or using isotype control antibody (Figure S8). We used ELISA to examine the expression of SFRP4 protein in DD-derived myofibroblasts. There was decreased accumulation of extracellular SFRP4 from the high-risk genotype cells (Figure 2E), consistent with the role of SFRP4 as a secreted WNT antagonist. As SFRP4 has been previously shown to bind to WNT3A, we first examined WNT3A expression by immunohistochemistry in our fixed surgical specimens and showed expression in surgically resected fibrotic DD tissue, but not in palm skin or using isotype control antibody (Figure S8). We then used recombinant human WNT3A, SFRP4, or a combination of the two proteins to stimulate DD-tissue-derived myofibroblasts, palmar-skin-derived fibroblasts, and non-palmar-skin-derived fibroblasts from four unrelated DD-affected individuals. We found no difference in expression of collagen type I or type III in any of the cells tested. WNT3A selectively increased the expression of α-SMA in the Dupuytren myofibroblasts. Furthermore, WNT3A increased signaling in both canonical and non-canonical WNT signaling pathways, as evidenced by increased AXIN2 (MIM: 604025) and CTGF (MIM: 121009) expression, respectively (Figure 2F). Interestingly, SFRP4 appeared to act as a selective antagonist of non-canonical signaling by WNT3A but had no significant effect on canonical signaling (Figure 2G).

Discussion
We have completed the largest GWAS to date in DD, the most common inherited disorder of connective tissue. Our results have almost tripled the known loci associated with this localized fibrosis and have also highlighted the role of fundamental biological processes in the pathophysiology of fibrosis, in the context of DD. Several associated loci harbor potentially attractive drug targets and are the subject of active further research. While we acknowledge that the mechanistic link between associated SNPs and pathophysiological function can often be obscure and requires experimental validation, we think that certain biological processes deserve discussion.

WNT Signaling
The importance of WNT signaling in fibrosis, as exemplified by DD, has been confirmed by this work. All previously reported loci that harbor WNT pathway genes have been replicated in this larger study, including WNT ligands WNT2, WNT4, and WNT7B, a co-signaling molecule RSPO2, and WNT antagonist SFRP4. Our detailed functional studies on the statistically most strongly associated variant (rs16879765) have suggested that a subtle imbalance of WNT signaling contributes to the fibrotic phenotype. We postulate that the decreased SFRP4 secretion seen in individuals homozygous for the high-risk allele at rs16879765 allows a subtle increase in WNT3A signaling through the non-canonical pathway. This could lead to greater α-SMA expression and hence contraction of the DD cords. This contraction is characteristic of the latter stages of DD and requires surgical treatment.

Taken as a whole, our genetic results suggest that subtle variations in the level of WNT signaling are likely to be responsible for the fibrosis seen in DD. The genetic variants cluster around ligands, a co-stimulatory molecule, and a WNT antagonist. This contrasts with variants in the WNT signaling pathway that predispose to cancer, which tend to be downstream of the receptor and lead to
receptor-independent signaling and unrestrained cellular growth and proliferation. While DD does share some clinical features with cancer, such as excess cellular proliferation, abnormal extracellular matrix deposition, and the tendency to recur after treatment, it is ultimately a benign phenotype.

Extracellular Matrix Modulation

Fibrotic disease is characterized by abnormal and excessive extracellular matrix (ECM) deposition. In DD, the abnormal fibrotic cords are composed mainly of collagen, with a higher type III to type I collagen ratio than in unaffected palmar fascia. Several of our additional associated loci harbor genes that are known to interact with and modulate the ECM: DDR2 (MIM: 191311) at chromosome 1q23.3 (rs17433710, OR = 0.81, \( p_{\text{meta}} = 1.99 \times 10^{-10} \)), MMP14 (MIM: 600754) at chromosome 14q11.2 (rs1042704, OR = 1.26, \( p_{\text{meta}} = 2.49 \times 10^{-19} \)), ITGA11 (MIM: 604789) at chromosome 15q23 (rs2306022, OR = 1.28, \( p_{\text{meta}} = 8.70 \times 10^{-11} \)), ACAN (MIM: 155760) at...
Discoidin domain receptor 2 (DDR2) is a membrane-bound receptor tyrosine kinase that contains an extracellular discoidin homology domain. The functional ligand for DDR2 is fibrillar collagen (types I–III), though it has also been shown to bind type X collagen. DDR2 has previously been shown to play a role in collagen production and migration through the basement membrane by skin fibroblasts. Furthermore, DDR2 plays a complex role in liver fibrosis. DDR2 expression is induced by acute liver injury in a mouse model, and expression of a constitutionally active form of DDR2 enhances proliferation and invasion of hepatic stellate cells. However, in contrast to the acute injury model, DDR2 knockout mice are more susceptible to chronic inflammation and fibrosis in a carbon tetrachloride model of chronic liver injury. Intriguingly, this increased susceptibility to chronic fibrosis is mediated in part by attenuating the interaction of hepatic stellate cells with macrophages, suggesting a link between DDR2 and pro-inflammatory pathways (see below).

DDR2 expression has also been shown to be increased both in mouse models of osteoarthritis (OA) and in human OA. In this context, the effect of DDR2 is mediated by its induction of matrix metalloproteinase 13 (MMP13), the major MMP responsible for type II collagen degradation in OA. Decreased expression of DDR2 in heterozygous knockout mice lead to the attenuation of OA after joint destabilization. This raises the possibility of cross-talk between DDR2 and MMP pathways that may be relevant in DD pathogenesis. DDR2 represents an attractive therapeutic target in DD and other fibrotic diseases and is currently under active investigation by several pharmaceutical companies.

Matrix metalloproteinase 14 (MMP14 or MT1-MMP) is a type 1 transmembrane protein and member of the MMP family of proteases, initially characterized for their ability to degrade the extracellular matrix. MMP14 was the first membrane-bound MMP to be discovered and was initially characterized as a pro-MMP2 activator, though now at least 42 substrates have been defined, including fibrillar collagen and the WNT antagonist DKK1. There is some evidence for the involvement of MMP14 in DD pathogenesis. In clinical trials, broad-spectrum MMP inhibition caused some individuals to develop DD. MMP14 is over-expressed in DD nodules, and knockdown of MMP14 in DD-derived cells reduced both contraction and MMP2 activation in vitro. Interestingly, knockdown of MMPs including MMP14 did not change the rate of collagen breakdown, suggesting that non-proteolytic effects of MMP14 are responsible for the pro-fibrotic phenotype. Further characterization of the mechanism of action of MMP14 in DD may lead to the validation of this protein as a therapeutic target in fibrosis.

ITGA11 encodes integrin alpha11, a member of the integrin family of cell-surface-adhesion receptors. These type I transmembrane proteins act as heterodimers composed of an alpha and beta subunit, and through binding to the ECM transmit both mechanical and chemical signals to the cell. Heterodimers consisting of integrin alpha11 beta1 are fibroblast-specific collagen receptors, which are mechanically induced, and regulate myofibroblast differentiation. Furthermore, ITGA11 has recently been shown to be overexpressed in lung samples from individuals with idiopathic pulmonary fibrosis. Also, a variant near PTK2 (encoding focal adhesion kinase [FAK] [MIM: 600758]) at chromosome 8q24.3 showed suggestive evidence of association in our discovery cohort (rs12677559, OR = 0.86, \( p = 5.75 \times 10^{-6} \); Table S3; Figure S9A), but we were unable to design a suitable assay within the replication set. FAK is a non-receptor tyrosine kinase that is an integral part of focal adhesion structure and is phosphorylated in response to integrin engagement. In a mouse model of hypertrophic scarring, another example of a localized fibrosis, fibroblast-specific FAK knockout attenuated both fibrosis and inflammation (see below), emphasizing the importance of this signaling pathway in fibrosis.

Inflammation
Chronic inflammation has long been recognized as a key player in the pathogenesis of fibrosis in multiple organs, including liver, kidney, lung, and heart. Recent work has highlighted the important role of inflammation in DD. In particular, tumor necrosis factor (TNF)—signaling via the WNT pathway—was demonstrated to selectively upregulate α-SMA and subsequent contractility in palmar-skin-derived fibroblasts from DD-affected individuals compared to unaffected control subjects. While our associated regions do not harbor many inflammatory genes, they do indicate how this cross-talk between TNF and WNT signaling might occur. MAP4K5 (MIM: 604923)—also known as germinal center kinase-related (GCKR)—at chromosome 14q22.1 (rs1032466, OR = 0.82, \( p_{meta} = 1.96 \times 10^{-17} \)) has been shown to be activated by both TNF and WNT3A, and decreased expression of MAP4K5 inhibits GSK3β phosphorylation and subsequent β-catenin accumulation in B lymphocytes. This suggests a key role for MAP4K5, integrating TNF and WNT signaling in DD fibrosis.

In conclusion, we have described the largest-scale GWAS to date in DD, a common disease that is a model human fibrotic condition. We discovered 17 additional variants predisposing to fibrosis, bringing the total described to 26. Analysis of heritability explained by these 26 variants compared to all common autosomal variants in our study suggests that there are many more common variants affecting predisposition to DD and that larger studies with greater power will detect further associated loci. We characterized the subtle nature of the genetic predisposition at our statistically most associated locus, thereby identifying a potential therapeutic target. In addition, our results have highlighted other specific biological pathways.
that are likely to play an important role in the pathogenesis of DD and in fibrosis more widely.

DD represents a human disease that is attractive for early-phase trials of experimental therapeutics, owing to the ready availability of affected individuals, ease of access to affected tissues, and the excision of fibrotic tissue as routine part of clinical care.

Accession Numbers

The datasets generated and analyzed during the current study are available in the European Genome-phenome Archive repository (EGA: EGAS00001001206 for BSSH-GODD, EGA: phs000303.v1.p1 for KORA, EGA: EGAS00000000043 for GODDAF). Information on how to access the UKHLS data can be found on the Understanding Society website (see Web Resources).

Supplemental Data

Supplemental Data include nine figures and three tables and can be found with this article online at http://dx.doi.org/10.1016/j.ajhg.2017.08.006.

Acknowledgments

We are grateful to all affected and control individuals who participated in the study. We thank Dr. William Rayner (Wellcome Trust Centre for Human Genetics, Oxford University) for the use of scripts and Dr. Hoffman (University of Freiburg) for ELISA reagents. We wish to thank the collaborators of the BSSH-GODD group, the GODDAF group, and the German Dupuytren Study Group for recruitment of case subjects, Dr. Mohammed R. Toliat (Köln) for support in genotype analysis, and Ramona Casper for technical assistance. D.F. is supported by an Intermediate Clinical Fellowship from the Wellcome Trust (097152/Z/11/Z). This work was supported by the NIHR Biomedical Research Centre, Oxford. K.B., P.N., and H.C.H. are supported by the Deutsche Forschungsgemeinschaft through the Cluster of Excellence on Cellular Stress Response in Aging-associated Diseases at the University of Cologne (EXC229) and the Köln Fortune Program of the Faculty of Medicine, University of Cologne. E.Z. and L.S. were supported by the Wellcome Trust (098051). The UK Household Longitudinal Study is led by the Institute for Social and Economic Research at the University of Essex and funded by the Economic and Social Research Council. The survey was conducted by NatCen and the genome-wide scan data were analyzed and deposited by the Wellcome Trust Sanger Institute. Information on how to access the UKHLS data can be found on the Understanding Society website (see Web Resources).

Received: June 26, 2017
Accepted: August 3, 2017
Published: September 7, 2017

Web Resources

1000 Genomes, http://www.internationalgenome.org/
European Genome-phenome Archive (EGA), https://www.ebi.ac.uk/ega
Genotyping Chips Strand and Build Files, http://www.well.ox.ac.uk/~wrayer/strand/

PLINK 1.9, https://www.cog-genomics.org/plink2/
R statistical software, http://www.r-project.org/
Understanding Society, https://www.understandingsociety.ac.uk/

References


