Identification of novel plasma proteomic biomarkers of Dupuytren Disease

Blake Hummer[1¶], Paola Sebastiani[2,3&], Anastasia Leshchyk[4&], Anastasia Gurinovich[2,3&], Cecilie Bager[5&], Morten Karsdal[5&], Signe Nielsen[5&], Charles Eaton[1*¶]

[1] Dupuytren Research Group, West Palm Beach, Florida, United States of America

[2] Institute for Clinical Research and Health Policy Studies, Tufts Medical Center, Boston, Massachusetts, United States of America

[3] Department of Medicine, Tufts University, Boston, Massachusetts, United States of America

[4] Department of Medicine, Computational Biomedicine Section, Chobanian & Avedisian School of Medicine, Boston University, Boston, Massachusetts, United States of America

[5] Nordic Bioscience, Herlev, Denmark

* Corresponding Author

Email: c.eaton.md@dupuytrens.org (CE)

[¶] These authors contributed equally to this work.

[&] These authors also contributed equally to this work.

Abstract

Dupuytren Disease (DD) is a chronic progressive disease that can cause disabling hand deformities. The most common treatments have high rates of complications and early recurrence. Dupuytren lacks a staging biomarker profile to develop preventive therapeutics to improve long-term outcomes. This multi-omic study aimed to create a DD blood proteomic biomarker profile by comparing DD plasma to a healthy control group. We measured circulating collagen metabolism peptides and found normal Collagen I synthesis but impaired Collagen I degradation in DD. We measured 6995 serum protein aptamers and identified 68 proteins with statistically significant differences from the control group. We developed two Diagnostic Proteomic Risk Scores (DPRS) based on hypothesis-free and hypothesis-based analyses. In independent data, our hypothesis-free and hypothesis-based DPRS distinguished Dupuytren from control subjects with 76.5% and 70.6% accuracy, respectively. Our hypothesis-based DPRS also distinguished DD subjects with different disease progression rates based on subject age at the time of their first corrective procedure ($p=0.0018$). This pilot study is the first to provide evidence that Collagen I accumulation in DD is due to impaired degradation rather than increased collagen synthesis. It also describes novel DPRS that have potential use as diagnostic and staging biomarker panels for Dupuytren disease.

Introduction

Dupuytren disease (DD) is a progressive fibrotic condition that often results in finger deformities called Dupuytren contracture (DC). DD affects an estimated 10 million Americans. DC is associated with loss of productivity, higher healthcare costs, and a greater prevalence of benign diseases[1].

The greatest need in DD care is preventing DC progression and recurrence. As with other fibrotic diseases, developing preventive treatments requires a biomarker profile for staging and measuring therapeutic response. The authors have not found published reports with statistically significant evidence of preventing DC recurrence or a biomarker tool to develop such treatments. This pilot study aimed to develop preliminary data on a diagnostic blood biomarker profile of DD.

DC is a common outcome of DD. Of those diagnosed with DD at age 65, one in four men and one in eight women will require at least one corrective procedure over their lifetime[2]. Current treatments are palliative, often with partial or temporary improvement[3]. Surgery for DC has a greater risk of major complications than other common elective hand operations[4-7], and recurrent DC is the most common diagnosis leading to elective finger amputation[8].

Genetic factors account for about 80% of the risk of developing DD[9]. Family history, age, and gender have strong associations. A genetic risk score (GRS) based on 11 SNPs correlates with DD diagnosis[10], and a weighted GRS based on 26 SNPs correlates with post-procedure recurrence[11].

DD affects the palm skin, subcutaneous tissues, and fascia[12]. Early disease is cellular, with fibroblasts, myofibroblasts, and perivascular clusters of myofibroblasts, pericytes, and mononuclear immune cells[13]. Early changes include local variations of immune[13],

hematopoietic, and mesenchymal cell markers[12], cytokines, signaling factors[14], and extracellular matrix components[15, 16]. Progression transforms tissues into nearly acellular, densely packed collagen.

Physical changes are the basis for DD diagnosis, staging, and treatment, but physical findings don't predict progression[17]. There are no standard noninvasive tests of DD biological activity. Biopsy is problematic because it can trigger disease activity and progression. In addition, DD-related tissue findings do not always correlate with blood measurements[18-20], and DD tissue gene expression does not correlate with local protein levels[21]. Blood remains the most promising source for a disease activity profile.

We conducted a discovery study to compare levels of circulating proteins and peptides in subjects with DD versus controls using ELISA, aptamer, and mass spectrometry methods. We found statistically significant differences in disease versus control cohorts in independent analyses of peptides and proteins. We also found statistically significant differences in protein aptamers in DD subjects with different rates of clinical progression of untreated disease.

Materials and methods

Sex as a biological variable

Our study involved male and female subjects. Sex was not considered a biological variable.

Subjects

We selected disease cohort subjects from the International Dupuytren Database (IDDB), a patient-reported registry developed by the Dupuytren Research Group (DRG) in collaboration with the National Databank of Rheumatic Diseases (NDB). Control subjects were enrolled by

word of mouth and from the NDB database (https://dupuytrens.org/enroll-in-the- iddb-2/). All subjects in the IDDB met the entry criteria and signed an Informed Consent Form.

Entry criteria

The DD cohort met criteria based on clinical evidence of DD with strong traditional predictors of recurrent deformity after a corrective procedure: age of disease onset younger than 50; two or more first-degree relatives with DD; current DD nodules; three or more separate locations of DD or two palm locations and Ledderhose disease (a related disorder). Control cohort subjects were healthy individuals with no clinical evidence of DD, no personal history of DD or Ledderhose disease, and no family history of DD.

We confirmed study eligibility by requiring candidates to upload recent images of each hand's palmar and dorsal sides. The Principal Investigator/ senior author confirmed entry criteria and reviewed images to confirm clinical evidence of active DD in the disease cohort and the lack of these findings in the control cohort.

Blood Sample Collection

The study used a mobile phlebotomy service to collect blood specimens (Phlebotek Solutions Corporation, Fort Lauderdale, FL). Once confirmed that the subject met all study entry criteria, they were contacted by DRG and later by Phlebotek to schedule their blood draw. The phlebotomist visited the subjects in their homes and collected two tubes of blood. The last 16 blood draws included a third 2.5ml PAXgene tube collection for additional genomic testing.

Processing of the EDTA tubes for proteomic testing was standardized. Immediately following collection, blood samples were centrifuged for 15 minutes at 2000 g to generate plasma. The plasma was harvested from the top of the container and pooled before aliquoting into storage

cryotubes. The last 500 µL of plasma was discarded to avoid contamination with platelets. Samples were immediately packaged for shipping on dry ice.

Number of subjects

One hundred subjects were planned (50 DD/50 Control). Blood collection launched just as the COVID pandemic began (i.e., February 2020). Due to COVID-related issues, we amended the protocol to enroll 50 subjects. Forty-seven subjects were enrolled;46 subjects had samples drawn. One sample was disqualified due to failed shipping refrigeration. Samples from the remaining 45 subjects (25 DD, 20 Control) had mass-spectrometry (Bioproximity, Manassas, VA). Twenty-seven subjects (13 DD, 14 control) had second blood samples drawn six to 14 months after their first draw for mass-spectrometry testing by a second lab (MacCoss Lab, Seattle, WA). Additional proteomic testing was conducted on remaining aliquots: 1) n= 27 redrawn samples had ELISA tests for collagen metabolites (Nordic Bioscience, Herlev, DK); 2) n= 72 from both initial and redraw specimens had Aptamer-based testing (SomaLogic, Boulder, CO). Supplementary Fig. S1 outlines the distribution of samples and laboratories.

Blinding

All subjects were assigned and identified by participant numbers. Investigators were blinded to the sample group allocation. All samples remained blinded throughout the processing and data collection in the laboratories.

Assays

Nordic Bioscience collagen markers

Circulating collagen synthesis and degradation markers were measured using proprietary assays[22]. Synthesis biomarkers included collagens I (PRO-C1), II (PRO-C2), III (PRO-C3), IV

126 (PRO-C4), V (PRO-C5), VI (PRO-C6), VII (PRO-C7), and IX (PRO-C9). Circulating degradation
127 markers were measured for collagens I (C1M), II (CALC2, C2M), III (C3M, CTX-III), IV (C4M), VI
128 (C6M), and VII (C7M). All biomarkers were assessed by competitive ELISAs and validated to
129 measure in human plasma samples. The inter- and intra-assay coefficients of variation were
130 >15% and 10%, respectively, for all assays.

131 Mass spectrometry

132 A novel untargeted proteomics assay using a large-scale liquid chromatography-tandem mass
133 spectrometry (LC-MS/MS) approach called data-independent acquisition (DIA) was used to
134 collect proteomic data. Each sample (10 µl plasma), following tryptic digestion, was diluted
135 50:50 by the addition of 4% sodium dodecyl sulfate (SDS) buffer to a final concentration of ~2%
136 SDS. Approximately 200 µg of protein was added to 96-well plates, and the remainder of the
137 sample preparation was performed using the single-pot, solid-phase enhanced sample
138 preparation for proteomics methodology (SP3) [23, 24]. The entire process of protein capture,
139 reduction and alkylation, protein cleanup, digestion, and peptide cleanup was performed using a
140 KingFisher magnetic bead robot[24]. Peptides were reconstituted in HPLC starting aqueous
141 buffer (5% acetonitrile and 0.1% formic acid) with analysis performed on the high-resolution
142 quadrupole-orbitrap hybrid mass spectrometer (Exploris 480, Orbitrap Fusion Lumos, or
143 Orbitrap Fusion Eclipse).

144 Data was acquired in a single, 2-hour LC-MS/MS run (90 min HPLC gradient) on a Thermo
145 Quadrupole-Orbitrap mass spectrometer. Crucially, the DIA method comprehensively samples
146 all peptides between 400 and 1000 m/z using a repeated cycle of 75 MS/MS scans with an 8
147 m/z-wide isolation width. Using a multiplexing scheme based on staggering the isolation window
148 in alternating cycles of MS/MS scans, the precursor selectivity of the resulting data is 4 m/z after
149 computational demultiplexing[25]. From the resulting data, fragment ion chromatograms were

150 extracted for any peptide precursor between 400 and 1000 m/z. To determine which peptides
151 are confidently detected in each acquisition, the software tool EncyclopeDIA[26] was used to
152 query the data using a reference chromatogram library generated from a prior analysis of the
153 tissue. Chromatogram-based relative quantitation was then performed for each peptide
154 detected. All measurements are made relative to a common reference. In the case of plasma, a
155 common reference is a pooled standard from 400 age-matched samples from subjects who are
156 both healthy and span the phenotypes being studied, processed in each sample batch. Within
157 each 96-well plate, one copy of the common reference standard is placed in each row for a total
158 of 8 copies per plate[27]. Before quantification, it is confirmed that peptides are stable during the
159 time scale of the data collection, have a suitable linear response, and have suitable intra- and
160 inter-data precision.

161 SomaScan

162 SomaLogic, Inc. (Boulder, CO, USA) performed proteomic analysis of the plasma samples using
163 a SomaScan platform that quantifies the presence of 6995 human proteins (secreted proteins,
164 extracellular domains, and intracellular proteins). These proteins are found in diverse biological
165 groups such as receptors, kinases, cytokines, proteolytic enzymes and inhibitors, growth
166 factors, protease inhibitors, hormones, and structural proteins[28]. Most of these proteins
167 involve pathways relating to signal transduction, stress response, immune processes,
168 phosphorylation, proteolysis, cell adhesion, cell differentiation, and intracellular transport[29].

169 Proteins were measured using the Slow Off-Rate Modified Aptamer (SOMAmer)-based capture
170 array that uses chemically modified single-stranded DNA sequences capable of uniquely
171 recognizing individual proteins via one or more binding sites. SOMAmers transform a protein
172 signal to a nucleotide signal that is quantified using relative fluorescence on microarrays. Each
173 SOMAmer array underwent validation for specificity, upper and lower limits of detection, and

both intra- and interassay variability. Plasma dilutions (0.005%, 0.05%, and 20%) were applied to capture low-, medium-, and high-abundance proteins. Positive and negative controls were also positioned on the array to confirm if the experimental procedure was performed correctly.

After the hybridization step, the microarrays were washed and scanned using a laser scanner that excites the fluorescence of the fluorochrome used in the labeling step. The amount of the emitted signal is directly proportional to the amount of dye on the microarray. The scanner measures this quantity and creates a digital image map of the position of each signal based on the location of its origin. Image analysis software generates a text file from this map describing the pixel intensity of the spot and the background. This information was processed, and a final value was generated that summarized the level of each detectable protein on the microarray.

Quality control analysis of the sample and SOMAmer levels involved SOMAmer control arrays on the microarray and calibration samples. Hybridization controls measured the sample-by-sample variation in hybridization, while the median signal over all SOMAmer arrays measured the technical variability. The SomaScan measures were reported as relative fluorescence units (RFU) in a summary ADAT file.

Analysis and statistics

Nordic collagen marker analysis

Nordic collagen data turnover rates were calculated by performing ratios of the propeptide over the mature peptide (e.g., PRO-C1/ C1M). This ratio is an estimate of the formation/degradation of each collagen type and can be interpreted as a net gain of collagen/loss of collagen remodeling[22]. All comparisons were done using Student's T-test.

SomaScan analysis

A series of quality control and data cleaning steps was performed on the SomaScan proteomics data. All non-human proteins were excluded from the analyses. Samples that did not pass the default normalization acceptance criteria for all row scale factors were removed. All aptamers that did not pass the default quality control criteria across all plates/sets were filtered out. The data had no missing values, and the protein values were log-transformed. No further normalization was performed.

We identified differentially expressed aptamers using linear regression of each log-transformed protein as the outcome and disease status as the main predictor. We adjusted each analysis by age and sex. We analyzed all repeated measurements of each individual and employed generalized estimating equations to account for technical and biological variability using the gee package in R. Due to the high number of aptamers tested in the SomaScan assay (6995), we were at a statistical testing disadvantage regarding multiple corrections. Therefore, we calculated adjusted p-values using Benjamini-Hochberg methods. Our SomaScan analyses only report proteins with statistically significant differences from control at an FDR-adjusted p-value < 0.25.

Mass spectrometry analysis

For the MS protein data, all zeroes were replaced with NAs across samples, and protein values were log-transformed. The MS dataset was subset based on proteins that overlapped with the signature proteins associated with DD previously identified using the SomaScan proteomics data. We used a generalized estimation equation model to account for the repeated blood draws, adjusted by age and gender.

Proteomic Risk Score calculations

To prepare the data for the Diagnostic Proteomic Risk Score (DPRS) calculation, we first removed all highly correlated proteins (Pearson correlation > 0.8). 3369 out of 6995 aptamers were highly correlated and were excluded from the DPRS calculation. The DPRS calculation contained the following two steps. First, we fit a logistic regression for each protein with disease status (i.e., case or control) as the outcome variable adjusted by age and gender. We used a generalized estimation equation model for repeated measurements of two independent blood draws. We included the same subject's repeated measurements and used the generalized estimation equations model with the gee R package[30] to estimate the regression coefficients. Next, the DPRS was computed using summary statistics from each regression model. Only the proteins with statistically significant differences from control at an FDR-adjusted p-value < 0.25 were included in the score calculation. The DPRS was calculated by summing the beta estimated effect for all significant proteins, each multiplied by normalized protein abundance and divided by its standard error. Finally, we fit a logistic regression to examine the association of the computed DPRS with DD onset. We split the dataset into training (75%) and testing sets utilizing the caret R package for a leave-one-out cross-validation approach for model fitting and assessment (25%) sets[31]. We used R version 4.2.1 for all analyses.

Pathway analysis

STRING (12.0) is a web tool that provides and illustrates protein-protein interactions and networks[32]. STRING integrates different types of evidence from public databases, including genomic context, high-throughput, and biochemical interaction experiments, known protein complexes, computational predictions, and text-mining from scientific literature. Given a list of gene names, STRING constructs protein-protein interaction networks where each gene is considered as a node, with edges representing the predicted functional associations.

Networks were generated using the whole genome as the statistical background, and interaction sources included text mining, experiments, databases, co-expression, neighborhood, gene fusions, and co-occurrence. Interaction scores were created by combining the probabilities from all evidence channels and correcting for the probability of randomly observing an interaction. All nodes and interactions were computed with a required medium confidence interaction score (0.400) and an FDR<=0.05.

Study Approval

The Institutional Review Board Ascension Via Christi Hospitals, Wichita, Inc., approved this study and the associated informed consent form. The written informed consent was signed and received before subject participation. This written informed consent included a description of the use of the photographs/ images to be reviewed for study entry. The records of these signed informed consents have been retained. All samples remained blinded throughout the processing and data collection in the laboratories. All methods were performed in accordance with relevant guidelines, regulations, and the Declaration of Helsinki.

Results

Plasma collagen metabolites in DD

Collagen synthesis and degradation processes release protein fragments specific to the process, collagen type, and enzyme involved. The ratio of synthesis to degradation-related fragments correlates with net accumulation or loss of that collagen type. We evaluated plasma collagen synthesis and degradation markers using the competitive ELISA method (Nordic Biosciences; Herlev, Denmark) and calculated collagen synthesis-to-degradation ratios[22].

262 We measured markers of types I, II, III, IV, V, VI, VII, and IX collagen metabolism in blood from a
263 cohort of healthy controls (n = 14) and subjects with DD (n = 13). We found significantly low
264 levels of markers of Collagen I degradation C1M (Figure 1A, p = 0.004), Collagen V synthesis
265 Pro-C5 (Figure 1B, p = 0.046), and Collagen VII synthesis Pro-C7 (Figure 1C, p = 0.048) in the
266 DD cohort compared to the control group. The remaining collagen biomarkers did not show
267 significant differences between the disease and control groups (Table 1).

268 We calculated collagen synthesis-to-degradation ratios for collagen types I, II, III, IV, and V. The
269 Collagen I synthesis to degradation ratio, Pro-C1/C1M, was significantly greater in the DD
270 cohort compared to the control group (Figure 1D, p = 0.024). The remaining ratios were not
271 significant (Col II, III, IV, VI) or could not be calculated because the degradation marker was
272 either unavailable (Col V, IX) or not detectable in both disease and control cohorts (Col VII).

273 Differentially expressed blood proteins in DD

274 We performed two orthogonal unbiased proteomic methods, the aptamer-based SomaScan
275 (SomaLogic; Boulder, CO) and liquid chromatography-tandem mass spectrometry (LC-MS/MS)
276 (MacCoss Lab, University of Washington, WA, USA).

277 We used sixty-nine samples for SomaScan testing after quality control (36 DD and 33 controls).
278 These represent 42 subjects, 27 of whom had second samples obtained six to fourteen months
279 after the first (Supplementary Fig. S1). The SomaScan panel measures 6995 protein binding
280 sites applying to 5444 proteins.

281 Hypothesis-free approach

282 The 6995 aptamer measurements identified 54 proteins with statistically significant differences
283 between DD and control groups, with 24 overexpressed and 30 underexpressed after adjusting

for multiple comparisons (Supplementary Table S1). Figure 2 is a heatmap of these 54 protein expressions.

We performed pathway analysis of these 54 proteins to identify network relationships using STRING[32]. We found significant protein-protein interactions (PPI) in 29, with multiple enriched categories (Figure 3C, p-value = 4.55E-04). Supplementary Table S2 summarizes categories of enriched gene groups, and Supplementary Fig. S2 provides a visual cross-reference of functional categories and individual genes. The dominant categories were coagulation and complement cascades, the extracellular matrix, enzyme inhibition, cell membrane structures, phosphorylation, and signaling pathways.

Using the SomaScan results, we aimed to develop a hypothesis-free indicator of DD diagnosis with a calculated Diagnostic Proteomic Risk Score (DPRS) to estimate an individual's disease risk based on plasma protein abundance.

Hypothesis-free DPRS

We developed a hypothesis-free DPRS estimation based on two of the 54 proteins. These proteins, OSCAR (overexpressed) and SPART (underexpressed), had a low relative correlation (Pearson's correlation coefficient = -0.19) (Table 2), suggesting minimal interaction. Using this hypothesis-free DPRS as a diagnostic measure for DD achieved an accuracy of 79% in the training set (AUC 0.8) and 76.5% in the test set (Figure 3A), indicating a predictive value for discriminating between disease and control. Figure 3B shows a box plot of control and DD DPRS. The DPRS calculated for controls and DD significantly differed by the Wilcoxon test (p-value = 2.5E-06).

Hypothesis-based approach

306 We also conducted a hypothesis-based approach with 328 relevant proteins identified from a
307 literature search and available on the SomaScan platform (Supplementary Table S3). Twenty-
308 three proteins in this subset had statistically significant differences between DD and control, with
309 nine overexpressed and fourteen underexpressed after adjusting for multiple comparisons
310 (Supplementary Table S4).

311 We performed PPI networks and enrichment analysis on these 23 proteins using STRING. We
312 found significantly enriched PPI in 21 proteins with multiple enriched categories (Figure 4C, p-
313 value = 1.35e-07). We found more enrichment categories and greater significance in the
314 hypothesis-based than the hypothesis-free analysis, which may reflect a higher-yield group,
315 selection bias, or both. Supplementary Table S5 lists enriched gene categories, and
316 Supplementary Fig. S3 provides a visual cross-reference between functional categories and
317 individual genes.

318 Hypothesis-based DPRS

319 We developed a hypothesis-based DPRS using 11 of 23 DE proteins identified in the
320 hypothesis-based analysis. Six genes were underexpressed (*SMAD1, SH3BP2, LCN2, USP8,*
321 *CSNK1G2, SERPINH1*) and five overexpressed (*CSMD1, POSTN, ACAN, AOC3, DDR2*). Six
322 proteins were present in hypothesis-based and hypothesis-free DPRS analyses (Table 3). Our
323 hypothesis-based DPRS achieved a training accuracy of 78.85% (AUC 0.9) and a test accuracy
324 of 70.59% in distinguishing Dupuytren from control cohorts (Figure 4A). Figure 4B shows that
325 DPRS distributions of Control and DD groups differ significantly (Wilcoxon test (p-value = 2.2E-
326 09).

327 Evaluation of DRPS for prognosis

We explored the potential of each DPRS to profile DD prognosis using the rate of contracture progression as an index of biological severity. We segregated our DD samples into subjects younger (11 specimens) or older (16 specimens) than 50 at the time of their first DC procedure. We excluded subjects younger than 50 from this analysis. A one-way ANOVA analysis unveiled a statistically significant variance in the means of hypothesis-driven generated DPRS across groups with different rates of pretreatment contracture progression (Supplementary Fig. S4, p-value = 0.0018). Conversely, we found no associations in the DPRS generated through the hypothesis-free approach.

Mass Spectrometry

The LC-MS dataset contained 48 samples: 27 DD and 21 controls. This assay identified 561 peptides. No peptides had statistically significant differences in abundance across groups (FDR > 0.25).

Mass-Spectrometry-SomaScan concordance

To further validate our SomaScan findings, we evaluated the MS results for concordance. Of the 54 significant proteins from the SomaScan data analysis, six were present in the Mass Spectrometry findings: AOC3, POSTN, SERPINC1, TF, PCSK9, and KNG1. We ran the same analyses on these six proteins using Mass Spectrometry data. AOC3, POSTN, and SERPINC1 had adjusted p-values (FDR) <0.25 in both Mass Spec and SomaScan analyses (Table 4).

Discussion

This study aimed to discover candidate biomarkers that correlate with the diagnosis of DD as the first step toward a diagnostic DD blood test.

349 Dynamic biomarker panels are essential to developing preventive therapeutics for chronic
350 diseases like DD. The relevance of this study is that DD lacks such biomarkers. Blood is the
351 most promising unexplored source of proteomic biomarkers for trend analysis.

352 There are no broad surveys of circulating proteins in DD. In addition, blood findings may not
353 parallel tissue findings because of rapid degradation, intracellular or membrane-bound
354 locations, post-translational variants, protein binding, and other mechanisms. Because of this,
355 new and diverse discovery research is necessary to profile DD-associated blood protein
356 biomarkers.

357 There are two common biomarker discovery strategies: hypothesis-free and hypothesis-based.
358 The hypothesis-free approach (unbiased or non-hypothesis-based) screens many markers and
359 is the standard in systems biology and discovery research. The hypothesis-based strategy
360 evaluates selected markers based on published information and is the most common model of
361 DD research. We used both methods in this study.

362 Plasma collagen metabolites in DD

363 Collagen accumulation is the hallmark of DD. DD-affected tissues have unique proportions of
364 collagen types I[33], III[34], IV[16], and V[33]. The most prominent tissue changes in the
365 progression from preclinical to end-stage disease are increases in total Collagen I and the
366 proportion of Collagen III[33].

367 Our hypothesis-based approach compared collagen synthesis and degradation markers in DD
368 versus control groups. Collagen turnover releases circulating peptides unique to the synthesis
369 and degradation of each collagen type. In the absence of disease, synthesis and degradation
370 markers are tightly coupled. Plasma synthesis to degradation marker ratios can indicate net
371 collagen gain, loss, or homeostasis[22]. We used ELISA to measure plasma levels of sixteen

372 collagen metabolism markers involving eight collagen types. We summarize our findings in
373 Table 1.

374 The DD cohort had normal Collagen I synthesis marker Pro-C1 levels but statistically significant
375 decreases in Collagen I degradation marker C1M (Figure 1A), resulting in an elevated
376 synthesis-to-degradation marker ratio of DD compared to controls (Figure 1D). These findings
377 suggest DD collagen accumulation follows impaired Collagen I degradation rather than excess
378 collagen synthesis. Both degradation-resistant collagen variants and reduced collagenase
379 activity may contribute to reduced collagen degradation.

380 Dupuytren collagen has two features that increase resistance to enzymatic degradation. Typical
381 Collagen I fibrils are heterotrimers of two Collagen I α1 strands and one Collagen I α2 strand.
382 DD-affected tissues have an increased proportion of α1 (I) homotrimer Collagen I[35]. DD
383 collagen also has more crosslinks stabilized by hydroxylation and glycosylation[36]. Both
384 homotrimer Collagen I[37] and glycosylation-stabilized crosslinks[38] increase resistance to
385 enzymatic cleavage.

386 DD-related dysregulation of collagenases and their inhibitors impairs collagen degradation by
387 several mechanisms. TIMP1 (Tissue Inhibitor of Metalloproteinase 1) inhibits multiple MMPs
388 and is elevated in the circulation of DD patients[18]. MMP1, MMP8, MMP13, and MMP14 gene
389 polymorphisms are associated with Dupuytren disease[39, 40], including MMP14 partial loss-of-
390 function variants[40]. MMP14 is necessary for MMP2 activation, and both enzymes facilitate
391 myofibroblast contraction[41].

392 Two clinical findings support a relationship between impaired collagen degradation and DD. The
393 first is the rapid onset of DD with metalloproteinase inhibitor treatment[42]. The second is
394 abnormally slow matrix remodeling in DD. Enzymatic cleavage of collagen fibrils is necessary
395 for matrix shortening[43], and impaired cleavage would slow this process. Slow matrix

396 remodeling could explain why DD-related tissue shortening progresses more slowly than wound
397 healing despite similar biology and why DD surgery has an increased risk of delayed wound
398 healing[44]. The role of impaired collagen degradation in pathogenesis and intervention is
399 unclear, and simply normalizing collagen degradation might have the undesirable effect of
400 accelerating tissue contraction.

401 The hallmark of fibrotic diseases is collagen accumulation. Increased collagen synthesis
402 contributes to collagen accumulation in pulmonary fibrosis[45], liver fibrosis[46], primary biliary
403 cirrhosis[47], renal fibrosis[48], and cardiovascular disease[49]. Our finding of normal levels of
404 circulating Collagen I synthesis biomarkers in DD suggests that the mechanism of collagen
405 accumulation in DD is not due to the increased collagen synthesis associated with these
406 diseases. Although often used interchangeably with increased collagen synthesis, neither
407 collagen accumulation nor upregulated collagen gene expression is de facto evidence of
408 increased collagen synthesis. Cell culture studies show increased *COL1A1* and *COL1A2* mRNA
409 expression in DD-derived fibroblasts[21, 50], but these RNA findings are not associated with
410 increased collagen I protein production[21]. Our finding of normal Collagen I synthesis and
411 impaired degradation informs future research because DD drug development has focused on
412 reducing collagen synthesis rather than increasing degradation[51].

413 Collagen II markers were normal. We evaluated Collagen II markers because of the correlation
414 between DD and psoriasis[52] and the finding of abnormal Collagen II in psoriasis[53]. We
415 found no significant differences between groups in Collagen II synthesis marker PRO-C2_HP
416 and degradation marker C2M_HP or their ratios. The CALC2 synthesis marker was
417 undetectable in both cohorts, preventing the calculation of CALC2 synthesis-to-degradation
418 ratios.

419 Collagen III synthesis markers were normal, consistent with a prior report of normal circulating
420 Collagen III synthesis marker amino-terminal propeptide of type III procollagen (PIIINP) in DD
421 subjects[54]. Given the accumulation of Collagen III in DD, we expected to find elevated
422 Collagen III synthesis to degradation marker ratios. Instead, we found normal Collagen III
423 synthesis to degradation ratios. This finding may reflect that other proteases also degrade
424 Collagen III, and we did not measure all potential collagen degradation fragments. To
425 investigate further, we assayed a different Col III degradation marker, CTX-III_HP, on the
426 remaining aliquots. Only 17 (10 DD, 7 Controls) of the initial 27 (13 DD, 14 Controls) leftover
427 aliquots had sufficient residual volumes for this assay. Although this small subgroup's CTX-III
428 violin plot appearance suggested reduced degradation in DD compared to controls
429 (Supplementary Fig. S5), this difference was not statistically significant.

430 We found significantly lower Collagen V synthesis marker Pro-C5 levels in the DD cohort than in
431 controls (Figure 1B). This finding is unexplained. An increased percentage of Collagen V is
432 reported in DD tissues[33], suggesting a high Collagen V synthesis to degradation ratio.
433 However, we could not calculate this ratio because a Collagen V degradation marker was
434 unavailable.

435 The DD cohort also had significantly lower Collagen VII synthesis marker Pro-C7 than controls
436 (Figure 1C). We could not calculate Collagen VII synthesis to degradation ratios because
437 Collagen VII degradation markers were undetectable in both cohorts. Although there are no
438 reports of Collagen VII in DD, its role in matrix mechanobiology and elevated tissue levels in
439 systemic sclerosis[55] make it a potential marker for collagen metabolism abnormalities in DD.

440 We found no significant differences in the measurements of the remaining collagen markers. We
441 measured Collagen IV markers because Collagen IV is increased in DD tissues[16]. We
442 assayed Collagen VI markers and ratios because prior pathway analysis suggests a role for

443 Collagen VI in DD and other fibrotic diseases[56]. We selected Collagen IX markers based on
444 unpublished data suggesting a role in DD. None of these markers or ratios showed significant
445 differences between disease and control cohorts.

446 Plasma proteins and interactions in DD

447 Separate from the Nordic collagen fragment analysis, we analyzed whole proteins with the
448 SomaScan aptamer-based platform. The SomaScan panel measures 6995 protein binding sites
449 associated with 5444 proteins. Supplementary Table S6 lists genes and gene aliases of proteins
450 mentioned in this manuscript.

451 SomaScanc hypothesis-free analysis identified 54 distinct proteins with significantly different
452 plasma levels than controls. Twenty-four proteins had increased levels, and thirty had lower DD
453 levels than controls (Supplementary Table S1). Our findings are consistent with existing
454 literature.

455 The body of DD research publications mentions only seven of these 54 differentially expressed
456 proteins: Aggrecan (*ACAN*), Amine Oxidase Copper Containing 3 (*AOC3*), Caspase-3 (*CASP3*),
457 Periostin (*POSTN*), SMAD Family Member 1 (*SMAD1*), Spartin (*SPART*), and Ubiquitin
458 carboxyl-terminal hydrolase 8 (*USP8*). Our differential expression findings concur with reports of
459 *ACAN*[57], *AOC3*[13], and *POSTN*[58] overexpression in tissue and *SMAD1*[59], *SPART*[60],
460 and *USP8*[60] underexpression in cell culture. CASP3 levels were low in our findings, but are
461 upregulated in a report of transcriptomic profiling[57]. This finding is unexplained. *ACAN* is also
462 near the variant rs6496519 associated with DD[61]. Supplementary Table S1 summarizes the
463 relevance of these 54 genes to DD.

464 We constructed PPI networks and enrichment analysis of these 54 plasma proteins, revealing
465 significantly enriched protein-protein interactions in 32, with multiple enriched protein groups
466 and categories (Supplementary Fig. S2, Supplementary Table S2). These categories are

467 aligned with known Dupuytren cellular biology. Abnormalities in the KEGG category has04610
468 (complement and coagulation cascades) were consistent with DD pathway meta-analysis
469 findings[62]. Complement fragments induce adhesion molecules that act on endothelial cells,
470 consistent with DD-related perivascular clusters of immune cells and small vessel occlusion[13].
471 Complement fragments also activate mast cells that release TNF, a known DD pathway
472 component[63], and histamine, a driver of itching, which is common in early-stage DD[64]. Cell
473 membrane and extracellular matrix signaling receptor binding of growth factors, cytokines, and
474 other signaling factors play pivotal roles in DD pathobiology[14]. Protein phosphorylation
475 processes are consistent with increased phosphorylated STAT1, STAT3, SMAD2, SMAD3, and
476 ERK1/2 in DD tissues[59, 65, 66]. The enriched Reactome HSA-162582 (Signal Transduction)
477 pathway is noteworthy, as it includes signaling by TGF-Beta, WNT, and NOTCH, all prominent
478 in DD biology[67].

479 We based our hypothesis-free DPRS on high levels of Osteoclast-associated immunoglobulin-
480 like receptor (*OSCAR*) and low levels of Spartin (*SPART*) (Table 2). Our finding of elevated
481 OSCAR levels is interesting because Collagen I is an OSCAR ligand. OSCAR also stimulates T-
482 cells to release TNF[68], a component of DD biology[63]. Our finding of decreased *SPART*
483 levels is consistent with low *SPART* expression in DD fibroblasts[60]. SPART is also associated
484 with Epidermal Growth Factor Receptor (EGFR) degradation and transport to the cell
485 membrane[69], consistent with increased ratios of intracellular to cell membrane EGFR levels
486 found in DD[70].

487 Separate from the hypothesis-free analysis, we conducted a hypothesis-based analysis using a
488 subset of the SomaScan proteins selected from an independent literature review
489 (Supplementary Table S3). We found 23 proteins in this subset with significantly different
490 plasma levels than controls. Nine proteins had higher, and fourteen had lower levels in DD than
491 in controls. Nine proteins were present in both hypothesis-free and hypothesis-based groups.

492 Our findings are consistent with existing literature. Supplementary Table S4 summarizes the
493 potential relationships of these proteins and parent genes to DD.

494 With our hypothesis-free analysis, we performed a STRING PPI analysis of the hypothesis-
495 based findings to identify network relationships of the 23 proteins. We found significantly
496 enriched protein-protein interactions in 21 with multiple enriched categories (p-value = 1.35e-07,
497 Supplementary Table S5). This analysis implicated enrichment categories in addition to those
498 from the hypothesis-free analysis. Myofibroblast differentiation, proliferation, motility, WNT
499 signaling, and interleukin pathways are unsurprising, given their known roles in DD biology[10,
500 20]. Cancer pathway associations are also unsurprising, given the correlations of DC with
501 carcinomas and sarcomas[71, 72]. Infection-related pathways were unexpected and exclusively
502 involved viral and parasitic infection categories. This finding might be due to a hidden variable of
503 tissue repair biology, which overlaps interleukin-driven parasitic immune defense biology and
504 interleukin-related DD biology.

505 This hypothesis-based analysis provided data to generate a second Diagnostic Proteomic Risk
506 Score (DPRS) calculated on eleven proteins (Table 3). Figure 4A shows the AUC-ROC of this
507 panel (AUC = 0.9), and Figure 4B shows the DPRS based on this panel, distinguishing DD from
508 control cohorts.

509 We evaluated the potential of each DPRS for staging. Recurrence after a procedure is the
510 benchmark of Dupuytren's biological severity. Traditionally, surgeons use clinical factors such
511 as early age of onset, affected family members, and disease locations other than the palms to
512 predict post-procedure recurrence. However, the study's inclusion criteria included these
513 factors, so we could not use them to subset our Dupuytren cohort. Instead, we used age at the
514 time of the first DC procedure. Small studies suggest that recurrence is more common in
515 patients younger at their first DC procedure[73, 74]. We evaluated the ability of each DPRS to
516 distinguish Dupuytren subjects who had their first corrective procedure before age 50 versus

517 those having their first correction at age 50 or older. The hypothesis-based DPRS distinguished
518 DD subjects requiring their first DC procedure before age 50 from those having their first
519 procedure after age 50 (p=0.0018), suggesting potential use as a staging and predictive tool.

520 Although Mass Spectrometry did not reveal statistically significant differences in abundance
521 across cohorts, there was statistically significant concordance between Mass Spec and
522 SomaScan findings. The Mass Spec data contained 6 of the 54 proteins with statistically
523 significant differences between disease and control cohorts using the hypothesis-free
524 SomaScan data analysis: Amine Oxidase Copper Containing 3 *AOC3*, Periostin *POSTN*, Serpin
525 Family C Member 1 *SERPINC1*, Transferrin *TF*, Proprotein Convertase Subtilisin/Kexin Type 9
526 *PCSK9*, and Kininogen 1 *KNG1*. Table 4 compares the results of the SomaScan and Mass
527 Spectrometry data analysis. A pathway analysis of these six proteins identified two highly
528 significant Reactome categories (Supplementary Table S7). In addition to protein
529 phosphorylation noted above, Reactome HSA-381426 involving Insulin-like Growth Factor *IGF*
530 and Insulin-like Growth Factor Binding Proteins *IGFBP*s is relevant because of the role of IGF-II
531 and IGFBP-6 in regulating Dupuytren fibroblast proliferation and contractility[75].

532 This study has limitations. Statistical disadvantages were the small number of samples in all
533 assays and the large number of aptamers in the SomaScan assay. Although our findings align
534 with current research on Dupuytren proteomics, there are differences. Elevated serum
535 TIMP1[18] and TNF[20] have been reported in DD, but we found no statistically significant DE
536 between disease and control groups. This discrepancy may reflect that SomaScan assays
537 target specific protein conformations, whereas ELISA measures all conformations.

538 Conclusions

539 We used ELISA, aptamer, and mass spectrometry assays to identify differential plasma
540 proteomic signatures between DD and control cohorts. We found significant DD-related

541 proteomic abnormalities in both collagen metabolism peptide fragments and whole proteins in
542 circulation. This study is the first to have evidence suggesting that net collagen accumulation in
543 DD derives from normal collagen synthesis and impaired collagen degradation, challenging the
544 model of excess collagen synthesis and distinguishing DD from other organ fibroses. We
545 independently analyzed proteins from the same cohorts. Our hypothesis-free analysis of 5444
546 proteins revealed 54 with statistically significant differences between the DD and control
547 cohorts. Our independent hypothesis-based analysis of 328 proteins selected from the literature
548 review identified 23 with statistically significant differences. We developed two Dupuytren
549 Diagnostic Proteomic Risk Scores separately based on hypothesis-free and hypothesis-driven
550 analyses. We used pathway analyses to identify DD protein-protein network enrichments,
551 confirming known DD interactions and revealing previously unreported DD pathway categories.
552 Finally, we showed our hypothesis-based DPRS correlated with the subject's age at the time of
553 the first DD procedure, an index of biological severity. These findings suggest that a validated
554 hypothesis-based DPRS could guide management and develop DD therapeutics.

555 We recommend a larger study to validate our Collagen I, III, V, and VII metabolite findings. We
556 also recommend a validating study of 68 unique proteins identified in the protein aptamer
557 analysis (54 in the hypothesis-free analysis and 23 in the hypothesis-based analysis, with 9 in
558 both groups). Testing only the subset of the 68 most pertinent proteins would increase the
559 statistical power. Individual ELISA assays would be cost-prohibitive. An alternative approach
560 would be to develop a custom multiplex panel for the 13 proteins in our two DPRS[14] to
561 validate our DPRS and the correlation of our hypothesis-based DPRS on age at the first DD
562 procedure. Our findings strongly support the need to fund and conduct more extensive
563 validating studies of circulating collagen metabolites and proteins in DD. This biomarker
564 research is fundamental to developing DD diagnostic tests to diagnose, stage, and inform
565 research to develop therapeutic drugs for DD.

Acknowledgments

Gail Solomon assisted with manuscript writing.

PS, AL, AG: recipients of grant UM1TR004398 (the Tufts NIH CTSA award).

Databank and study support from FORWARD, the National Databank for Rheumatic Diseases, is gratefully acknowledged.

Author Contributions

CE provided the study design and wrote the main manuscript text. CE and BH wrote the results section, and BS and PS wrote the methods section. All authors performed data analysis.

References

1. Macaulay D, Ivanova J, Birnbaum H, Sorg R, Skodny P. Direct and indirect costs associated with Dupuytren's contracture. J Med Econ. 2012;15(4):664-71. Epub 2012/03/01. doi: 10.3111/13696998.2012.670678. PubMed PMID: 22369346.

2. Broekstra DC, Kuo RYL, Burn E, Prieto-Alhambra D, Furniss D. Dupuytren Disease: Prevalence, Incidence, and Lifetime Risk of Surgical Intervention. A Population-Based Cohort Analysis. Plast Reconstr Surg. 2023;151(3):581-91. Epub 20221122. doi: 10.1097/PRS.0000000000009919. PubMed PMID: 36730480; PubMed Central PMCID: PMCPMC9944385.

3. Mendelaar NHA, Poelstra R, van Nieuwenhoven CA, Slijper HP, Feitz R, Hovius SER, et al. Outcome of Recurrent Surgery in Dupuytren's Disease: Comparison with Initial Treatment. Plast Reconstr Surg. 2019;144(5):828e-35e. Epub 2019/08/10. doi: 10.1097/PRS.0000000000006150. PubMed PMID: 31397794.

587 4. Goodman AD, Gil JA, Starr AM, Akelman E, Weiss A-PC. Incidence of Serious
588 Complications in Hand Surgery: A 10-Year Review. The Journal of Hand Surgery.
589 2016;41(9):S27-S8. doi: 10.1016/j.jhsa.2016.07.047.
590 5. Schindelar L, McEntee R, D'Amore T, Beredjiklian P, Lutsky K. Unplanned Return to the
591 Operating Room in Upper-Extremity Surgery: Incidence and Reason for Return. J Hand Surg
592 Am. 2021;46(8):715 e1- e12. Epub 20210514. doi: 10.1016/j.jhsa.2021.01.019. PubMed PMID:
593 33994259.
594 6. Denkler K. Surgical complications associated with fasciectomy for dupuytren's disease: a
595 20-year review of the English literature. Eplasty. 2010;Jan 27(10):e15. PubMed Central PMCID:
596 PMCPMC2828055.
597 7. Alser O, Craig RS, Lane JCE, Prats-Uribe A, Robinson DE, Rees JL, et al. Serious
598 complications and risk of re-operation after Dupuytren's disease surgery: a population-based
599 cohort study of 121,488 patients in England. Sci Rep. 2020;10(1):16520. Epub 2020/10/07. doi:
600 10.1038/s41598-020-73595-y. PubMed PMID: 33020582; PubMed Central PMCID:
601 PMCPMC7536429.
602 8. Degreef I, De Smet L. Dupuytren's disease: a predominant reason for elective finger
603 amputation in adults. Acta Chir Belg. 2009;109(4):494-7. Epub 2009/10/07. doi:
604 10.1080/00015458.2009.11680467. PubMed PMID: 19803262.
605 9. Larsen S, Krogsgaard DG, Aagaard Larsen L, Iachina M, Skytthe A, Frederiksen H.
606 Genetic and environmental influences in Dupuytren's disease: a study of 30,330 Danish twin
607 pairs. J Hand Surg Eur Vol. 2015;40(2):171-6. Epub 2014/05/20. doi:
608 10.1177/1753193414535720. PubMed PMID: 24835475; PubMed Central PMCID:
609 PMCPMC4810018.
610 10. Dolmans GH, Werker PM, Hennies HC, Furniss D, Festen EA, Franke L, et al. Wnt
611 signaling and Dupuytren's disease. N Engl J Med. 2011;365(4):307-17. Epub 2011/07/08. doi:
612 10.1056/NEJMoa1101029. PubMed PMID: 21732829.

613 11. Riesmeijer SA, Manley OWG, Ng M, Nolte IM, Broekstra DC, Werker PMN, et al. A
614 Weighted Genetic Risk Score Predicts Surgical Recurrence Independent of High-Risk Clinical
615 Features in Dupuytren's Disease. Plast Reconstr Surg. 2019;143(2):512-8. Epub 2019/01/29.
616 doi: 10.1097/PRS.0000000000005208. PubMed PMID: 30688894; PubMed Central PMCID:
617 PMCPMC6358194.
618 12. Iqbal SA, Hindocha S, Farhatullah S, Paus R, Bayat A. Dupuytren's Disease Shows
619 Populations of Hematopoietic and Mesenchymal Stem-Like Cells Involving Perinodular Fat and
620 Skin in Addition to Diseased Fascia: Implications for Pathogenesis and Therapy. Dupuytren's
621 Disease and Related Hyperproliferative Disorders2012. p. 167-74.
622 13. Mayerl C, Del Frari B, Parson W, Boeck G, Piza-Katzer H, Wick G, et al.
623 Characterisation of the inflammatory response in Dupuytren's disease. J Plast Surg Hand Surg.
624 2016;50(3):171-9. Epub 2016/02/09. doi: 10.3109/2000656X.2016.1140054. PubMed PMID:
625 26852784.
626 14. Gonga-Cave BC, Pena Diaz AM, O'Gorman DB. Biomimetic analyses of interactions
627 between macrophages and palmar fascia myofibroblasts derived from Dupuytren's disease
628 reveal distinct inflammatory cytokine responses. Wound Repair Regen. 2021;29(4):627-36.
629 Epub 20210611. doi: 10.1111/wrr.12928. PubMed PMID: 34212454.
630 15. Halliday NL, Rayan GM, Zardi L, Tomasek JJ. Distribution of ED-A and ED-B containing
631 fibronectin isoforms in Dupuytren's disease. J Hand Surg Am. 1994;19(3):428-34. Epub
632 1994/05/01. doi: 10.1016/0363-5023(94)90057-4. PubMed PMID: 8056970.
633 16. Magro G, Fraggetta F, Travali S, Lanzafame S. Immunohistochemical expression and
634 distribution of alpha2beta1, alpha6beta1, alpha5beta1 integrins and their extracellular ligands,
635 type IV collagen, laminin and fibronectin in palmar fibromatosis. Gen Diagn Pathol.
636 1997;143(4):203-8. Epub 1998/03/07. PubMed PMID: 9489951.
637 17. Broekstra DC, Lanting R, Werker PMN, van den Heuvel ER. Disease Course of Primary
638 Dupuytren Disease: 5-Year Results of a Prospective Cohort Study. Plast Reconstr Surg.

639 2022;149(6):1371-8. Epub 20220411. doi: 10.1097/PRS.0000000000009115. PubMed PMID:
640 35404338.
641 18. Ulrich D, Hrynyschyn K, Pallua N. Matrix metalloproteinases and tissue inhibitors of
642 metalloproteinases in sera and tissue of patients with Dupuytren's disease. Plast Reconstr Surg.
643 2003;112(5):1279-86. Epub 2003/09/25. doi: 10.1097/01.PRS.0000081462.40448.49. PubMed
644 PMID: 14504511.
645 19. Jupp O, Pullinger M, Marjoram T, Lott M, Chojnowski AJ, Clark IM. Biomarkers of
646 Postsurgical Outcome in Dupuytren Disease. Dupuytren Disease and Related Diseases - The
647 Cutting Edge2017. p. 55-61.
648 20. Wang ML, Rajpar I, Ruggiero NA, Fertala J, Steplewski A, Beredjiklian PK, et al.
649 Circulating inflammatory cytokines alter transcriptional activity within fibrotic tissue of
650 Dupuytren's disease patients. J Orthop Res. 2021. Epub 2021/04/30. doi: 10.1002/jor.25059.
651 PubMed PMID: 33913534.
652 21. van Beuge MM, Ten Dam EJ, Werker PM, Bank RA. Matrix and cell phenotype
653 differences in Dupuytren's disease. Fibrogenesis Tissue Repair. 2016;9:9. Epub 2016/07/02.
654 doi: 10.1186/s13069-016-0046-0. PubMed PMID: 27366208; PubMed Central PMCID:
655 PMCPMC4928329.
656 22. Karsdal MA, Leeming DJ, Henriksen K, Bay-Jensen AC, Nielsen SH, CL B. Biochemistry
657 of Collagens, Laminins and Elastin. Boston: Elsevier; 2019.
658 23. Hughes CS, Moggridge S, Muller T, Sorensen PH, Morin GB, Krijgsveld J. Single-pot,
659 solid-phase-enhanced sample preparation for proteomics experiments. Nat Protoc.
660 2019;14(1):68-85. Epub 2018/11/23. doi: 10.1038/s41596-018-0082-x. PubMed PMID:
661 30464214.
662 24. Leutert M, Rodriguez-Mias RA, Fukuda NK, Villen J. R2-P2 rapid-robotic
663 phosphoproteomics enables multidimensional cell signaling studies. Mol Syst Biol.

664 2019;15(12):e9021. doi: 10.15252/msb.20199021. PubMed PMID: 31885202; PubMed Central
665 PMCID: PMCPMC6920700.
666 25. Amodei D, Egertson J, MacLean BX, Johnson R, Merrihew GE, Keller A, et al. Improving
667 Precursor Selectivity in Data-Independent Acquisition Using Overlapping Windows. J Am Soc
668 Mass Spectrom. 2019;30(4):669-84. Epub 20190122. doi: 10.1007/s13361-018-2122-8.
669 PubMed PMID: 30671891; PubMed Central PMCID: PMCPMC6445824.
670 26. Searle BC, Pino LK, Egertson JD, Ting YS, Lawrence RT, MacLean BX, et al.
671 Chromatogram libraries improve peptide detection and quantification by data independent
672 acquisition mass spectrometry. Nat Commun. 2018;9(1):5128. Epub 2018/12/05. doi:
673 10.1038/s41467-018-07454-w. PubMed PMID: 30510204; PubMed Central PMCID:
674 PMCPMC6277451.
675 27. Hoofnagle AN, Becker JO, Oda MN, Cavigiolio G, Mayer P, Vaisar T. Multiple-reaction
676 monitoring-mass spectrometric assays can accurately measure the relative protein abundance
677 in complex mixtures. Clin Chem. 2012;58(4):777-81. Epub 20120203. doi:
678 10.1373/clinchem.2011.173856. PubMed PMID: 22307200; PubMed Central PMCID:
679 PMCPMC3665768.
680 28. Gold L, Ayers D, Bertino J, Bock C, Bock A, Brody EN, et al. Aptamer-based multiplexed
681 proteomic technology for biomarker discovery. PLoS One. 2010;5(12):e15004. Epub 20101207.
682 doi: 10.1371/journal.pone.0015004. PubMed PMID: 21165148; PubMed Central PMCID:
683 PMCPMC3000457.
684 29. Berrone E, Chiorino G, Guana F, Benedetti V, Palmitessa C, Gallo M, et al. SOMAscan
685 Proteomics Identifies Novel Plasma Proteins in Amyotrophic Lateral Sclerosis Patients. Int J Mol
686 Sci. 2023;24(3). Epub 20230118. doi: 10.3390/ijms24031899. PubMed PMID: 36768220;
687 PubMed Central PMCID: PMCPMC9916400.
688 30. Network TCRA. gee: Generalized Estimation Equation Solver 2023 [01/06/2023].
689 Available from: https://cran.r-project.org/package=gee.

690 31. Kuhn M. Building Predictive Models in R Using the caret Package. Journal of Statistical
691 Software [Internet]. 2008; 28(5).
692 32. Szklarczyk D, Kirsch R, Koutrouli M, Nastou K, Mehryary F, Hachilif R, et al. The
693 STRING database in 2023: protein-protein association networks and functional enrichment
694 analyses for any sequenced genome of interest. Nucleic Acids Res. 2023;51(D1):D638-D46.
695 doi: 10.1093/nar/gkac1000. PubMed PMID: 36370105; PubMed Central PMCID:
696 PMCPMC9825434.
697 33. Bailey AJ. Collagen. In: McFarlane R, McGrouther D, Flint M, editors. Dupuytren's
698 Disease Biology and Treatment. Hand and Upper Limb Series. 5. Edinburg: Churchill
699 Livingstone; 1990.
700 34. Bailey AJ, Sims TJ, Gabbiani G, Bazin S, LeLous M. Collagen of Dupuytren's disease.
701 Clin Sci Mol Med. 1977;53(5):499-502. Epub 1977/11/01. doi: 10.1042/cs0530499. PubMed
702 PMID: 589933.
703 35. Williamson K, Cooper G, Lee KJ, Beamish EL, Simpson D, Carter A, et al. Active
704 synthesis of type I collagen homotrimer in Dupuytren’s fibrosis is unaffected by anti-TNF-α
705 treatment. bioRxiv preprint 2023. doi: 10.1101/2020.07.13.195107.
706 36. Bazin S, Le Lous M, Duance VC, Sims TJ, Bailey AJ, Gabbiani G, et al. Biochemistry
707 and histology of the connective tissue of Dupuytren's disease lesions. Eur J Clin Invest.
708 1980;10(1):9-16. Epub 1980/02/01. doi: 10.1111/j.1365-2362.1980.tb00003.x. PubMed PMID:
709 6768572.
710 37. Han S, Makareeva E, Kuznetsova NV, DeRidder AM, Sutter MB, Losert W, et al.
711 Molecular mechanism of type I collagen homotrimer resistance to mammalian collagenases. J
712 Biol Chem. 2010;285(29):22276-81. Epub 2010/05/14. doi: 10.1074/jbc.M110.102079. PubMed
713 PMID: 20463013; PubMed Central PMCID: PMCPMC2903388.

714 38. Snedeker JG, Gautieri A. The role of collagen crosslinks in ageing and diabetes - the
715 good, the bad, and the ugly. Muscles Ligaments Tendons J. 2014;4(3):303-8. Epub 2014/12/10.
716 PubMed PMID: 25489547; PubMed Central PMCID: PMCPMC4241420.
717 39. Rodrigues MP, Tissi LH, Oliveira VM, Wistuba G, Araujo FB, Mattar-Junior R, et al.
718 MMP-1, MMP-8, and MMP-13 Gene Polymorphisms and Haplotype Is a Risk Factor for
719 Dupuytren Contracture: A Case-Control Study. Hand (N Y). 2024:15589447241242818. Epub
720 20240425. doi: 10.1177/15589447241242818. PubMed PMID: 38660990.
721 40. Itoh Y, Ng M, Wiberg A, Inoue K, Hirata N, Paiva KBS, et al. A common SNP risk variant
722 MT1-MMP causative for Dupuytren's disease has a specific defect in collagenolytic activity.
723 Matrix Biol. 2021;97:20-39. Epub 2021/02/17. doi: 10.1016/j.matbio.2021.02.003. PubMed
724 PMID: 33592276.
725 41. Wilkinson JM, Davidson RK, Swingler TE, Jones ER, Corps AN, Johnston P, et al. MMP-
726 14 and MMP-2 are key metalloproteases in Dupuytren's disease fibroblast-mediated
727 contraction. Biochim Biophys Acta. 2012;1822(6):897-905. Epub 2012/02/22. doi:
728 10.1016/j.bbadis.2012.02.001. PubMed PMID: 22342364.
729 42. Krzeski P, Buckland-Wright C, Balint G, Cline GA, Stoner K, Lyon R, et al. Development
730 of musculoskeletal toxicity without clear benefit after administration of PG-116800, a matrix
731 metalloproteinase inhibitor, to patients with knee osteoarthritis: a randomized, 12-month,
732 double-blind, placebo-controlled study. Arthritis Res Ther. 2007;9(5):R109. Epub 2007/10/26.
733 doi: 10.1186/ar2315. PubMed PMID: 17958901; PubMed Central PMCID: PMCPMC2212568.
734 43. Castella LF, Buscemi L, Godbout C, Meister JJ, Hinz B. A new lock-step mechanism of
735 matrix remodelling based on subcellular contractile events. J Cell Sci. 2010;123(Pt 10):1751-60.
736 Epub 2010/04/30. doi: 10.1242/jcs.066795. PubMed PMID: 20427321.
737 44. Bismil Q, Bismil M, Bismil A, Neathey J, Gadd J, Roberts S, et al. The development of
738 one-stop wide-awake dupuytren's fasciectomy service: a retrospective review. JRSM Short Rep.

739 2012;3(7):48. Epub 2012/08/22. doi: 10.1258/shorts.2012.012050. PubMed PMID: 22908029;
740 PubMed Central PMCID: PMCPMC3422854.
741 45. Organ LA, Duggan AR, Oballa E, Taggart SC, Simpson JK, Kang'ombe AR, et al.
742 Biomarkers of collagen synthesis predict progression in the PROFILE idiopathic pulmonary
743 fibrosis cohort. Respir Res. 2019;20(1):148. Epub 20190712. doi: 10.1186/s12931-019-1118-7.
744 PubMed PMID: 31299951; PubMed Central PMCID: PMCPMC6624898.
745 46. Liu T, Wang X, Karsdal MA, Leeming DJ, Genovese F. Molecular serum markers of liver
746 fibrosis. Biomark Insights. 2012;7:105-17. Epub 20120723. doi: 10.4137/BMI.S10009. PubMed
747 PMID: 22872786; PubMed Central PMCID: PMCPMC3412619.
748 47. Vesterhus M, Nielsen MJ, Hov JR, Saffioti F, Manon-Jensen T, Leeming DJ, et al.
749 Comprehensive assessment of ECM turnover using serum biomarkers establishes PBC as a
750 high-turnover autoimmune liver disease. JHEP Rep. 2021;3(1):100178. Epub 20200903. doi:
751 10.1016/j.jhepr.2020.100178. PubMed PMID: 33225252; PubMed Central PMCID:
752 PMCPMC7666353.
753 48. Han DC, Isono M, Hoffman BB, Ziyadeh FN. High glucose stimulates proliferation and
754 collagen type I synthesis in renal cortical fibroblasts: mediation by autocrine activation of TGF-
755 beta. J Am Soc Nephrol. 1999;10(9):1891-9. doi: 10.1681/ASN.V1091891. PubMed PMID:
756 10477140.
757 49. Genovese F, Goncalves I, Holm Nielsen S, Karsdal MA, Edsfeldt A, Nilsson J, et al.
758 Plasma levels of PRO-C3, a type III collagen synthesis marker, are associated with arterial
759 stiffness and increased risk of cardiovascular death. Atherosclerosis. 2024;388:117420. Epub
760 20231210. doi: 10.1016/j.atherosclerosis.2023.117420. PubMed PMID: 38128431.
761 50. Satish L, Gallo PH, Baratz ME, Johnson S, Kathju S. Reversal of TGF-beta1 stimulation
762 of alpha-smooth muscle actin and extracellular matrix components by cyclic AMP in
763 Dupuytren's-derived fibroblasts. BMC Musculoskelet Disord. 2011;12:113. Epub 2011/05/27.

764 doi: 10.1186/1471-2474-12-113. PubMed PMID: 21612641; PubMed Central PMCID:
765 PMCPMC3125251.
766 51. Lambi AG, Popoff SN, Benhaim P, Barbe MF. Pharmacotherapies in Dupuytren Disease:
767 Current and Novel Strategies. J Hand Surg Am. 2023. Epub 20230317. doi:
768 10.1016/j.jhsa.2023.02.003. PubMed PMID: 36935324.
769 52. Patel M, Freeman NR, Dhaliwal S, Wright N, Daoud Y, Ryan C, et al. The prevalence of
770 Dupuytren contractures in patients with psoriasis. Clin Exp Dermatol. 2014;39(8):894-9. Epub
771 2014/09/10. doi: 10.1111/ced.12424. PubMed PMID: 25196833.
772 53. Holm Nielsen S, Magee C, Groen SS, Sinkeviciute D, Bay-Jensen AC, Karsdal MA, et al.
773 Differentiating patients with psoriasis from psoriatic arthritis using collagen biomarkers. Clin Exp
774 Rheumatol. 2022. Epub 20220728. doi: 10.55563/clinexprheumatol/jmt9jv. PubMed PMID:
775 35916294.
776 54. Gassier N, Brenner P, Berger A, Delbruck A. Biochemical Parameters for the Diagnosis
777 of Dupuytren's Disease. In: Berger A, Delbruck A, Brenner P, Hinzmann B, editors. Dupuytren's
778 Disease Pathobiochemistry and Clinical Management. Berlin: Springer-Verlag; 1994. p. 94-8.
779 55. Sand JMB, Lamy P, Juhl P, Siebuhr AS, Iversen LV, Nawrocki A, et al. Development of a
780 Neo-Epitope Specific Assay for Serological Assessment of Type VII Collagen Turnover and Its
781 Relevance in Fibroproliferative Disorders. Assay Drug Dev Technol. 2018;16(2):123-31. Epub
782 20180301. doi: 10.1089/adt.2017.820. PubMed PMID: 29493258.
783 56. Williams LM, McCann FE, Cabrita MA, Layton T, Cribbs A, Knezevic B, et al. Identifying
784 collagen VI as a target of fibrotic diseases regulated by CREBBP/EP300. Proc Natl Acad Sci U
785 S A. 2020;117(34):20753-63. Epub 2020/08/08. doi: 10.1073/pnas.2004281117. PubMed PMID:
786 32759223; PubMed Central PMCID: PMCPMC7456151.
787 57. Rehman S, Salway F, Stanley JK, Ollier WE, Day P, Bayat A. Molecular phenotypic
788 descriptors of Dupuytren's disease defined using informatics analysis of the transcriptome. J

789 Hand Surg Am. 2008;33(3):359-72. Epub 2008/03/18. doi: 10.1016/j.jhsa.2007.11.010. PubMed
790 PMID: 18343292.
791 58. Vi L, Feng L, Zhu RD, Wu Y, Satish L, Gan BS, et al. Periostin differentially induces
792 proliferation, contraction and apoptosis of primary Dupuytren's disease and adjacent palmar
793 fascia cells. Exp Cell Res. 2009;315(20):3574-86. Epub 2009/07/22. doi:
794 10.1016/j.yexcr.2009.07.015. PubMed PMID: 19619531; PubMed Central PMCID:
795 PMCPMC5017872.
796 59. Krause C, Kloen P, Ten Dijke P. Elevated transforming growth factor beta and mitogen-
797 activated protein kinase pathways mediate fibrotic traits of Dupuytren's disease fibroblasts.
798 Fibrogenesis Tissue Repair. 2011;4(1):14. Epub 2011/06/30. doi: 10.1186/1755-1536-4-14.
799 PubMed PMID: 21711521; PubMed Central PMCID: PMCPMC3148569.
800 60. Forrester HB, Temple-Smith P, Ham S, de Kretser D, Southwick G, Sprung CN.
801 Genome-wide analysis using exon arrays demonstrates an important role for expression of
802 extra-cellular matrix, fibrotic control and tissue remodelling genes in Dupuytren's disease. PLoS
803 One. 2013;8(3):e59056. Epub 2013/04/05. doi: 10.1371/journal.pone.0059056. PubMed PMID:
804 23554969; PubMed Central PMCID: PMCPMC3595223.
805 61. Ng M, Thakkar D, Southam L, Werker P, Ophoff R, Becker K, et al. A Genome-wide
806 Association Study of Dupuytren Disease Reveals 17 Additional Variants Implicated in Fibrosis.
807 Am J Hum Genet. 2017;101(3):417-27. Epub 2017/09/09. doi: 10.1016/j.ajhg.2017.08.006.
808 PubMed PMID: 28886342; PubMed Central PMCID: PMCPMC5591021.
809 62. Park TH, Kim D, Lee YS, Kim SY. A meta-analysis to identify novel diagnostic and
810 therapeutic targets for Dupuytren's disease. Wound Repair Regen. 2020;28(2):202-10. Epub
811 2019/11/07. doi: 10.1111/wrr.12774. PubMed PMID: 31688987.
812 63. Verjee LS, Verhoekx JS, Chan JK, Krausgruber T, Nicolaidou V, Izadi D, et al. Unraveling
813 the signaling pathways promoting fibrosis in Dupuytren's disease reveals TNF as a therapeutic
814 target. Proc Natl Acad Sci U S A. 2013;110(10):E928-37. Epub 2013/02/23. doi:

815 10.1073/pnas.1301100110. PubMed PMID: 23431165; PubMed Central PMCID:
816 PMCPMC3593900.
817 64. Eaton C. Dupuytren Disease. In: Wolfe S, Pederson W, Kozin S, Cohen M, editors.
818 Green's Operative Hand Surgery. 1. Eighth ed. Philadelphia: Elsevier; 2022. p. 148-81.
819 65. Akbar M, Garcia-Melchor E, Chilaka S, Little KJ, Sood S, Reilly JH, et al. Attenuation of
820 Dupuytren's fibrosis via targeting of the STAT1 modulated IL-13Ralpha1 response. Sci Adv.
821 2020;6(28):eaaz8272. Epub 2020/07/23. doi: 10.1126/sciadv.aaz8272. PubMed PMID:
822 32695877; PubMed Central PMCID: PMCPMC7351483.
823 66. Kida H, Jiang JJ, Matsui Y, Takahashi I, Hasebe R, Kawamura D, et al. Dupuytren's
824 contracture-associated SNPs increase SFRP4 expression in nonimmune cells including
825 fibroblasts to enhance inflammation development. Int Immunol. 2023. Epub 20230131. doi:
826 10.1093/intimm/dxad004. PubMed PMID: 36719100.
827 67. Riesmeijer SA, Kamali Z, Ng M, Drichel D, Piersma B, Becker K, et al. A genome-wide
828 association meta-analysis implicates Hedgehog and Notch signaling in Dupuytren's disease.
829 Nat Commun. 2024;15(1):199. Epub 20240103. doi: 10.1038/s41467-023-44451-0. PubMed
830 PMID: 38172110; PubMed Central PMCID: PMCPMC10764787.
831 68. Schultz HS, Nitze LM, Zeuthen LH, Keller P, Gruhler A, Pass J, et al. Collagen induces
832 maturation of human monocyte-derived dendritic cells by signaling through osteoclast-
833 associated receptor. J Immunol. 2015;194(7):3169-79. Epub 20150227. doi:
834 10.4049/jimmunol.1402800. PubMed PMID: 25725106; PubMed Central PMCID:
835 PMCPMC4367160.
836 69. Bakowska JC, Jupille H, Fatheddin P, Puertollano R, Blackstone C. Troyer syndrome
837 protein spartin is mono-ubiquitinated and functions in EGF receptor trafficking. Mol Biol Cell.
838 2007;18(5):1683-92. Epub 20070301. doi: 10.1091/mbc.e06-09-0833. PubMed PMID:
839 17332501; PubMed Central PMCID: PMCPMC1855030.

840 70. Augoff K, Tabola R, Kula J, Gosk J, Rutowski R. Epidermal growth factor receptor (EGF-
841 R) in Dupuytren's disease. J Hand Surg Br. 2005;30(6):570-3. Epub 2005/08/02. doi:
842 10.1016/j.jhsb.2005.06.008. PubMed PMID: 16055243.
843 71. Wilbrand S. Dupuytren´s Contracture : Features and Consequences. Uppsala: Acta
844 Universitatis Upsaliensis; 2002.
845 72. Kuo RYL, Ng M, Prieto-Alhambra D, Furniss D. Dupuytren's Disease Predicts Increased
846 All-Cause and Cancer-Specific Mortality: Analysis of a Large Cohort from the U.K. Clinical
847 Practice Research Datalink. Plast Reconstr Surg. 2020;145(3):574e-82e. Epub 2020/02/26. doi:
848 10.1097/PRS.0000000000006551. PubMed PMID: 32097318; PubMed Central PMCID:
849 PMCPMC7043723.
850 73. van Rijssen AL, ter Linden H, Werker PM. Five-year results of a randomized clinical trial
851 on treatment in Dupuytren's disease: percutaneous needle fasciotomy versus limited
852 fasciectomy. Plast Reconstr Surg. 2012;129(2):469-77. Epub 2011/10/12. doi:
853 10.1097/PRS.0b013e31823aea95. PubMed PMID: 21987045.
854 74. Wilbrand S, Ekbom A, Gerdin B. The sex ratio and rate of reoperation for Dupuytren's
855 contracture in men and women. J Hand Surg Br. 1999;24(4):456-9. Epub 1999/09/03. doi:
856 10.1054/jhsb.1999.0154. PubMed PMID: 10473157.
857 75. Raykha C, Crawford J, Gan BS, Fu P, Bach LA, O'Gorman DB. IGF-II and IGFBP-6
858 regulate cellular contractility and proliferation in Dupuytren's disease. Biochim Biophys Acta.
859 2013;1832(10):1511-9. Epub 2013/04/30. doi: 10.1016/j.bbadis.2013.04.018. PubMed PMID:
860 23623986.

861 # Supporting information

862

863 ## Data availability

All the data and analytic scripts that support the findings of this study are openly available in the GitHub repository Dup_Blood_Proteomic_Data at https://github.com/Dupuytren/Dup_Blood_Proteomic_Data.

Competing interest statement

These authors declare the following competing interests: Morten Karsdal, Cecilie Bager, and Signe Holm Nielsen are all currently employed by Nordic Bioscience; Morten Karsdal, Cecilie Bager, and Signe Holm Nielsen hold stocks in Nordic Bioscience. The rest of the authors have no conflicts of interest to declare.

Figure sources

Figure 1. Each panel image was generated in R https://www.r-project.org. The final image was assembled and annotated with Moho 13.5.5 https://moho.lostmarble.com.

Figure 2. The image was generated in R https://www.r-project.org.

Figure 3. Panels a and b were generated in R https://www.r-project.org. Panel C was downloaded from STRING https://string-db.org. The final image was assembled and annotated with Moho 13.5.5 https://moho.lostmarble.com.

Figure 4. Panels a and b were generated in R https://www.r-project.org. Panel C was downloaded from STRING https://string-db.org. The final image was assembled and annotated with Moho 13.5.5 https://moho.lostmarble.com.

Supplementary Fig. S1 was created in PowerPoint.

Supplementary Fig. S2 was created in Excel.

Supplementary Fig. S3 was created in Excel.

885 Supplementary Fig. S4 was generated in R https://www.r-project.org.

886 Supplementary Fig. S5 was generated in R https://www.r-project.org.

887 Figure legends

888 Figure 1. Differential production of Dupuytren vs. control collagen metabolism markers. A)
889 Collagen I degradation marker C1M. B) Collagen I synthesis to degradation marker ratio Pro-
890 C1/C1M. C) Collagen V synthesis marker Pro-C5. D) Collagen VII synthesis marker Pro-C7. * p-
891 value <0.05, **p-value <0.005.

892 Figure 2. Heatmap of log-transformed standardized values of the 54 significant proteins in the
893 hypothesis-free analysis.

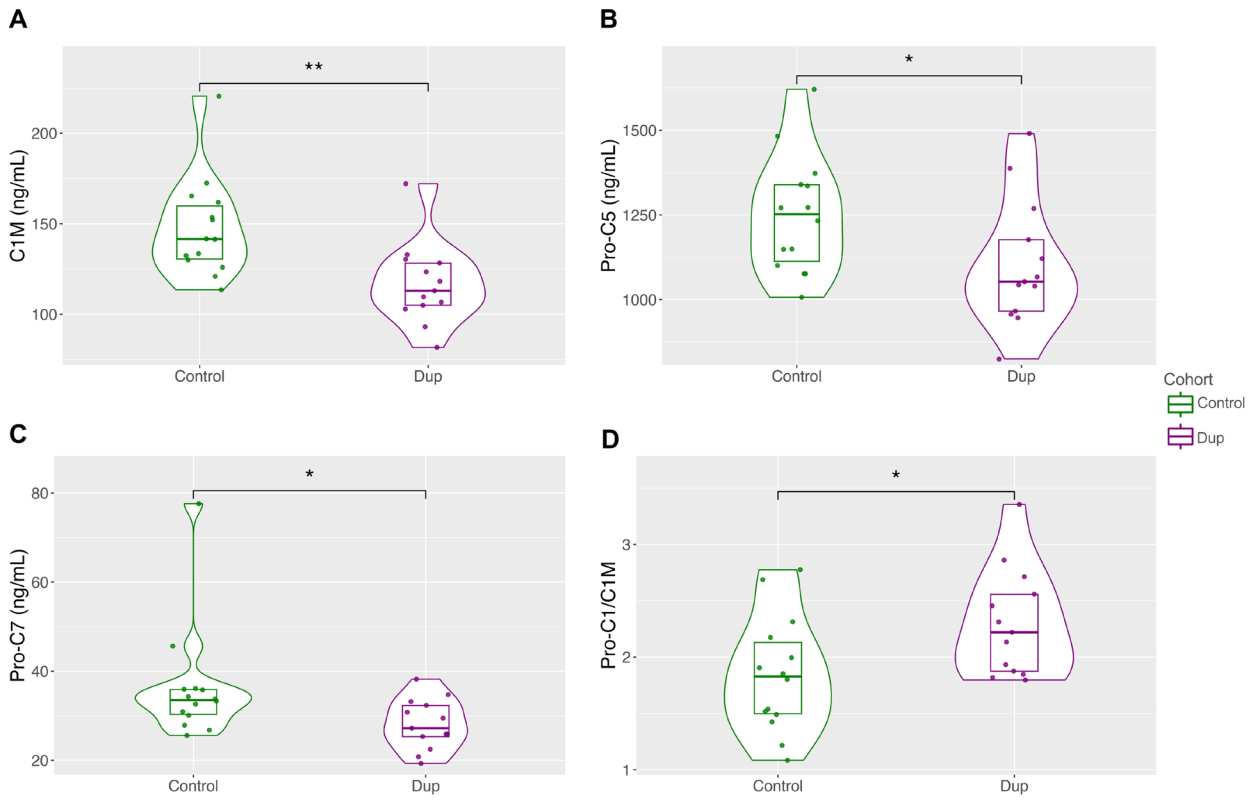
894 Figure 3. Hypothesis-Free DPRS and Protein-Protein Interactions. A) The Hypothesis-free
895 Diagnostic Proteomic Risk Score has an AUC-ROC value of 0.8. B) Hypothesis-free Diagnostic
896 Proteomic Risk Scores (HF-DPRS) of the control vs Dupuytren cohorts. p-value = 5.28E-06. C)
897 Network expression of 54 differentially expressed proteins. Of the 54 proteins, 30 had strong
898 pathway connections in the network: ACAN, AKR1A1, C5, C6, CALCB, CASP3, CGA, CGB3,
899 CPB1, CRKL, DSG2, EIF4H, FKBP5, G3BP2, GP1BB, GUSB, HGS, HSP90AA1, KNG1,
900 MAP3K11, MBNL1, MBNL2, NRG4, OSCAR, PCSK9, POSTN, PRSS1, SERPINC1, SOCS3,
901 SYK, TF, USP8, and YWHAB.

902

903 Figure 4. Hypothesis-based DPRS and Protein-protein interactions. A) The Hypothesis-based
904 Diagnostic Proteomic Risk Score has an AUC-ROC value of 0.9. AUC-ROC (Area Under the
905 Curve - Receiver Operating Characteristics) is an index of predictive accuracy. B) Hypothesis-
906 free Diagnostic Proteomic Risk Scores (HF-DPRS) of the control vs Dupuytren cohorts. (p-value
907 = 7.79E-09). C) Network expression of 23 differentially expressed proteins. Of the 23 proteins,
908 20 had strong pathway connections: DAB2, SMAD1, USP8, PLAT, SFRP4, ACAN, DDR2, HPX,

909 KNG1, LCN2, SERPINH1, TF, MAP2K1, PRKCA, STAT1, YWHAZ, POSTN, COL1A1, CASP3,
910 and STAT3.
911

912 Figure 1.



913

914

Figure 2.

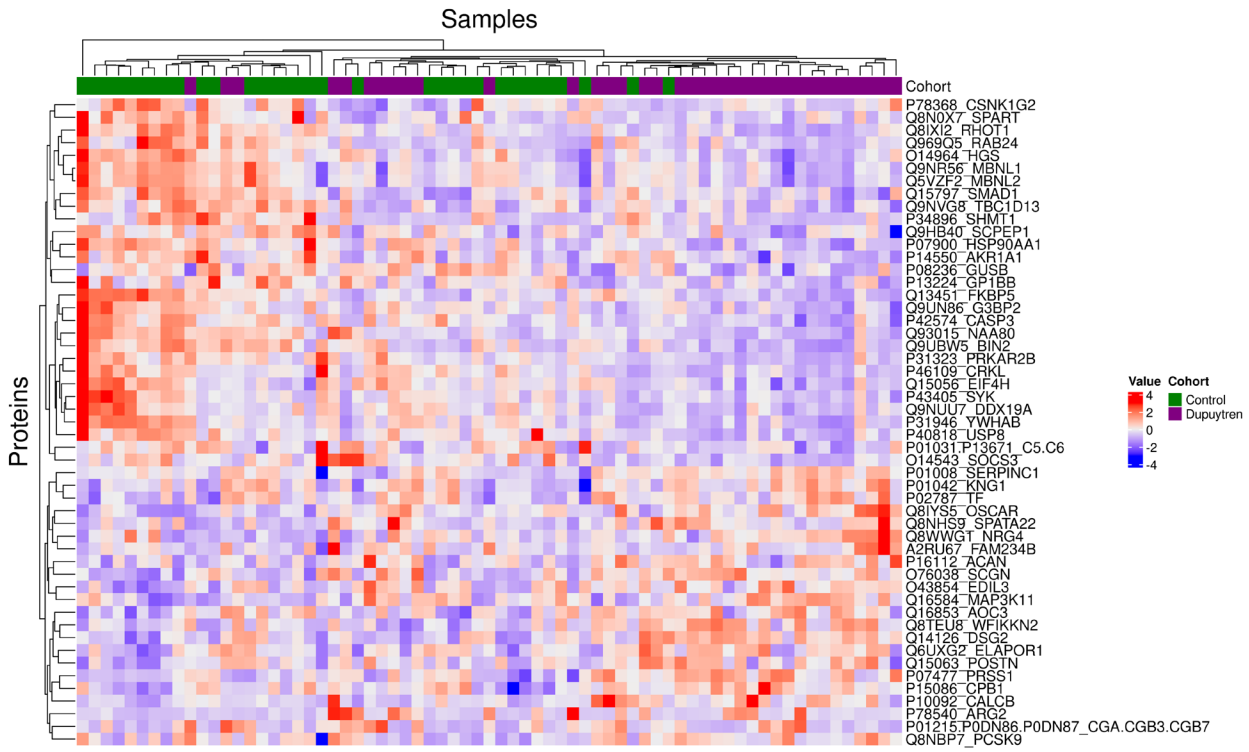
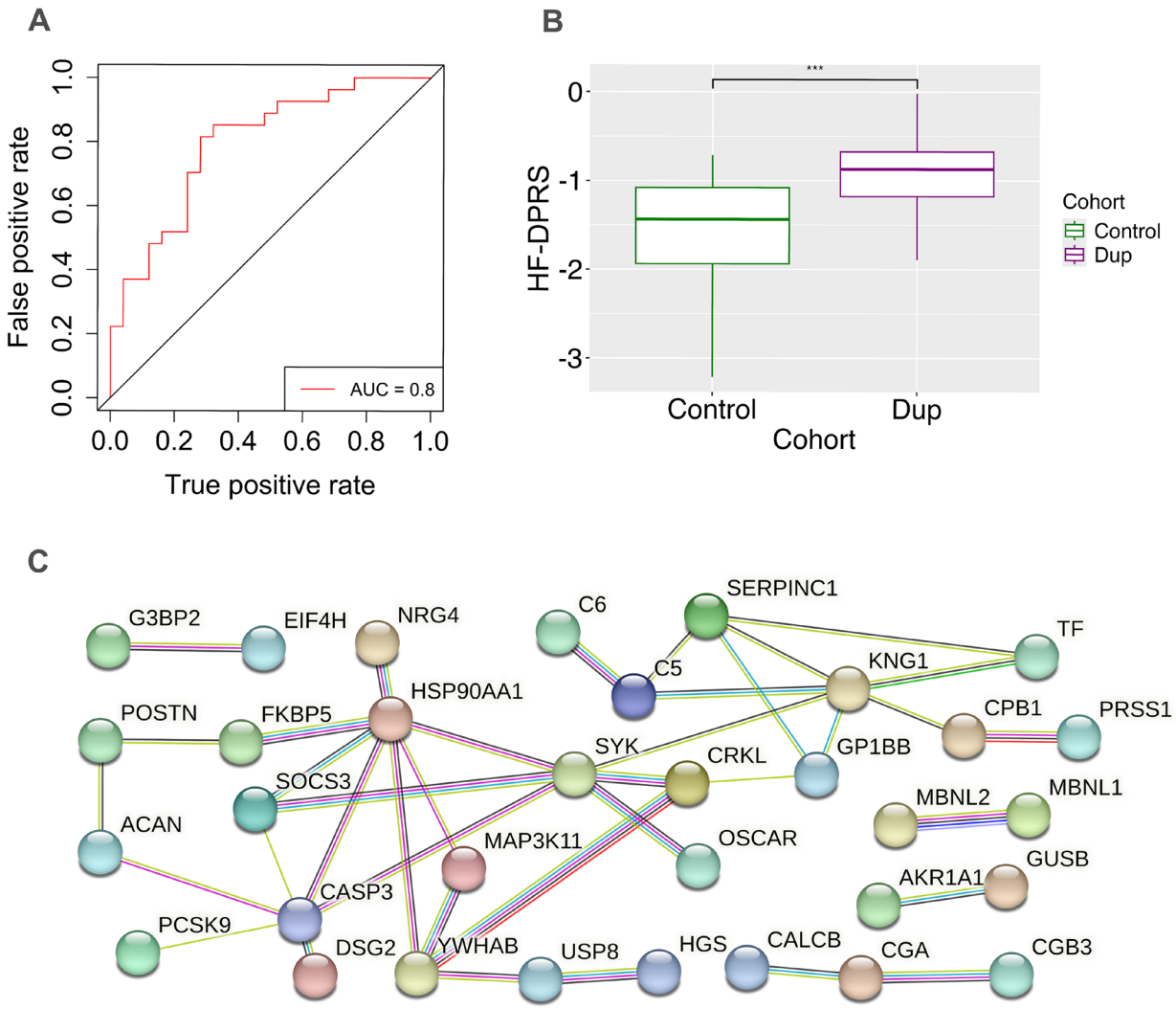
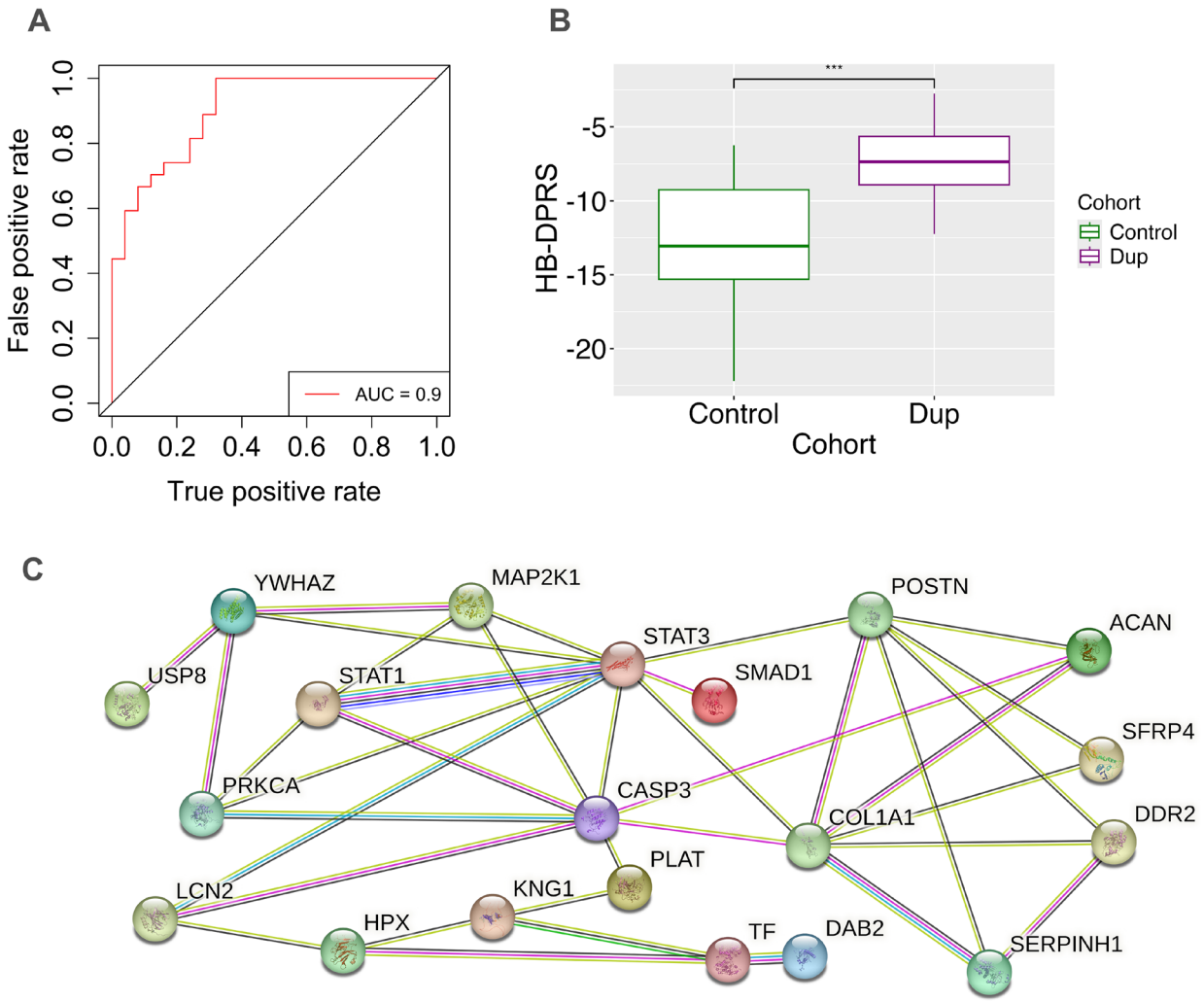

Samples
Cohort
Proteins
P78368_CSNK1G2
Q8N0X7_SPART
Q8IXI2_RHOT1
Q969Q5_RAB24
O14964_HGS
Q9NR56_MBNL1
Q5VZF2_MBNL2
Q15797_SMAD1
Q9NVG8_TBC1D13
P34896_SHMT1
Q9HB40_SCPEP1
P07900_HSP90AA1
P14550_AKR1A1
P08236_GUSB
P13224_GP1BB
Q13451_FKBP5
Q9UN86_G3BP2
P42574_CASP3
Q93015_NAA80
Q9UBW5_BIN2
P31323_PRKAR2B
P46109_CRKL
Q15056_EIF4H
P43405_SYK
Q9NUU7_DDX19A
P31946_YWHAB
P40818_USP8
P01031.P13671_C5.C6
O14543_SOCS3
P01008_SERPINC1
P01042_KNG1
P02787_TF
Q8IYS5_OSCAR
Q8NHS9_SPATA22
Q8WWG1_NRG4
A2RU67_FAM234B
P16112_ACAN
O76038_SCGN
O43854_EDIL3
Q16584_MAP3K11
Q16853_AOC3
Q8TEU8_WFIKKN2
Q14126_DSG2
Q6UXG2_ELAPOR1
Q15063_POSTN
P07477_PRSS1
P15086_CPB1
P10092_CALCB
P78540_ARG2
P01215.P0DN86.P0DN87_CGA.CGB3.CGB7
Q8NBP7_PCSK9
Value
4
2
0
-2
-4
Cohort
Control
Dupuytren

Figure 3.


A
False positive rate
True positive rate
0.0
0.2
0.4
0.6
0.8
1.0
AUC = 0.8
B
HF-DPRS
0
-1
-2
-3
***
Control
Dup
Cohort
Cohort
Control
Dup
C
G3BP2
EIF4H
NRG4
C6
SERPINC1
TF
C5
KNG1
HSP90AA1
POSTN
FKBP5
CPB1
PRSS1
SYK
CRKL
GP1BB
SOCS3
MBNL2
MBNL1
ACAN
MAP3K11
OSCAR
GUSB
AKR1A1
CASP3
PCSK9
DSG2
YWHAB
USP8
HGS
CALCB
CGA
CGB3

921 Figure 4.



922

923

Table 1. Collagen epitope markers and findings

Assay	D/C	Description	Represents	p-value	SE	FC (D/C)
PRO_C1_HP	ns	Internal epitope in the N-terminal propeptide of type I collagen	Col I Synthesis	0.9437	C 10.99 D 11.89	0.93
C1M_HP	Down**	MMP-generated fragment of the a1 chain of type I collagen	Col I Degradation	0.0036	C 5.25 D 4.30	0.73
PRO-C2_HP	ns	N-terminal fragment of type IIB procollagen	Col II Synthesis	0.7808	C 1.45 D 3.11	1.01
CALC2	ND	C-terminal propeptide of type II collagen	Col II Synthesis	NA	NA	NA
C2M_HP	ns	MMP-generated fragment of type II collagen	Col II Degradation	0.1607	C 1.52 D 1.16	1.29
PRO-C3_roHP	ns	N-terminal propeptide of type III collagen	Col III Synthesis	0.9951	C 1.04 D 1.24	0.93
C3M_HP	ns	MMP-9 generated fragments of type III collagen	Col III Degradation	0.1615	C 16.96 D 13.61	0.86
CTX-III_HP	ns	MMP-generated fragment of crosslinked type III collagen	Col III Degradation	0.4887	C 40.88 D 53.26	0.91
PRO-C4_HP	ns	Internal epitope in the 7S domain of type IV collagen	Col IV Synthesis	0.0739	C 195.98 D 183.12	0.84
C4M_HP	ns	MMP-generated fragment of the a1 chain of type IV collagen	Col IV Degradation	0.1014	C 8.89 D 7.42	0.84
PRO-C5_HP	Down*	C-terminal propeptide of type V collagen	Col V Synthesis	0.0458	C 33.17 D 35.88	0.82
PRO-C6_roHP	ns	C-terminal propeptide of type VI collagen alpha-3 chain	Col VI Synthesis	0.7222	C 0.27 D 0.27	0.92
C6M_HP	ns	MMP2-generated fragment of type VI collagen	Col VI Degradation	0.1097	C 22.31 D 23.41	0.83
PRO_C7 HP	Down*	BMP-I mediated neo-epitope in the NC-2 propeptide of type VII collagen	Col VII Synthesis	0.0480	C 2.49 D 1.08	0.72
C7M	ND	Neo-epitope of MMP13 degradation of type VII collagen	Col VII Degradation	NA	NA	NA
PRO-C9	ns	C-terminus of NC-I domain of the a1 chain of type IX collagen	Col IX Synthesis	0.9658	C 0.80 D 0.65	0.92

Synthesis and degradation markers of Collagen I, II, III, IV, V, VII, and IX. Compared to controls, there were significant reductions in markers of Collagen I degradation, Collagen V synthesis, and Collagen VII synthesis. D/C: ratio of Dupuytren to control values. FC: fold change. **: <0.005, *: <0.05. ns: not significant. ND: undetectable in both disease and control cohorts.

924

925

Table 2. Hypothesis-free Diagnostic Proteomic Risk Score

Protein	Eff	SE	p-value	FDR
Spartin	-11.62125	2.97479	0.000094	0.18
Oscar	13.99037	3.596067	0.0001	0.18

Proteins of the Hypothesis-free DPRS. Eff = log (odds ratio comparing Dup to control) for a change of 1 in the log (protein data). SE: Standard error. FDR: False Discovery Rate (Adjusted p-value).

926

927

Table 3. Hypothesis-based Diagnostic Proteomic Risk Score

Protein	Eff	SE	p-value	FDR
SMAD1*	-6.3649	1.877804	0.0007	0.12
CSMD1	5.397535	1.907423	0.00466	0.18
SH3BP2	-4.59151	1.582341	0.00371	0.18
POSTN*	4.305245	1.502627	0.00417	0.18
ACAN*	5.589134	1.75746	0.00147	0.12
LCN2	-4.73895	1.734879	0.0063	0.19
AOC3*	7.050175	2.62144	0.00716	0.2
DDR2	1.63418	0.595212	0.00604	0.19
USP8*	-3.30095	1.041269	0.00152	0.12
CSNK1G2*	-10.3331	3.593633	0.00404	0.18
SERPINH1	-2.6813	0.799538	0.0008	0.12

Proteins of the Hypothesis-based DPRS. Eff = log (odds ratio comparing Dup to control) for a change of 1 in the log (protein data). SE: Standard error. FDR: False Discovery Rate (Adjusted p-value). * Five also had significant differential expression in the 54 proteins identified in the hypothesis-free analysis.

928

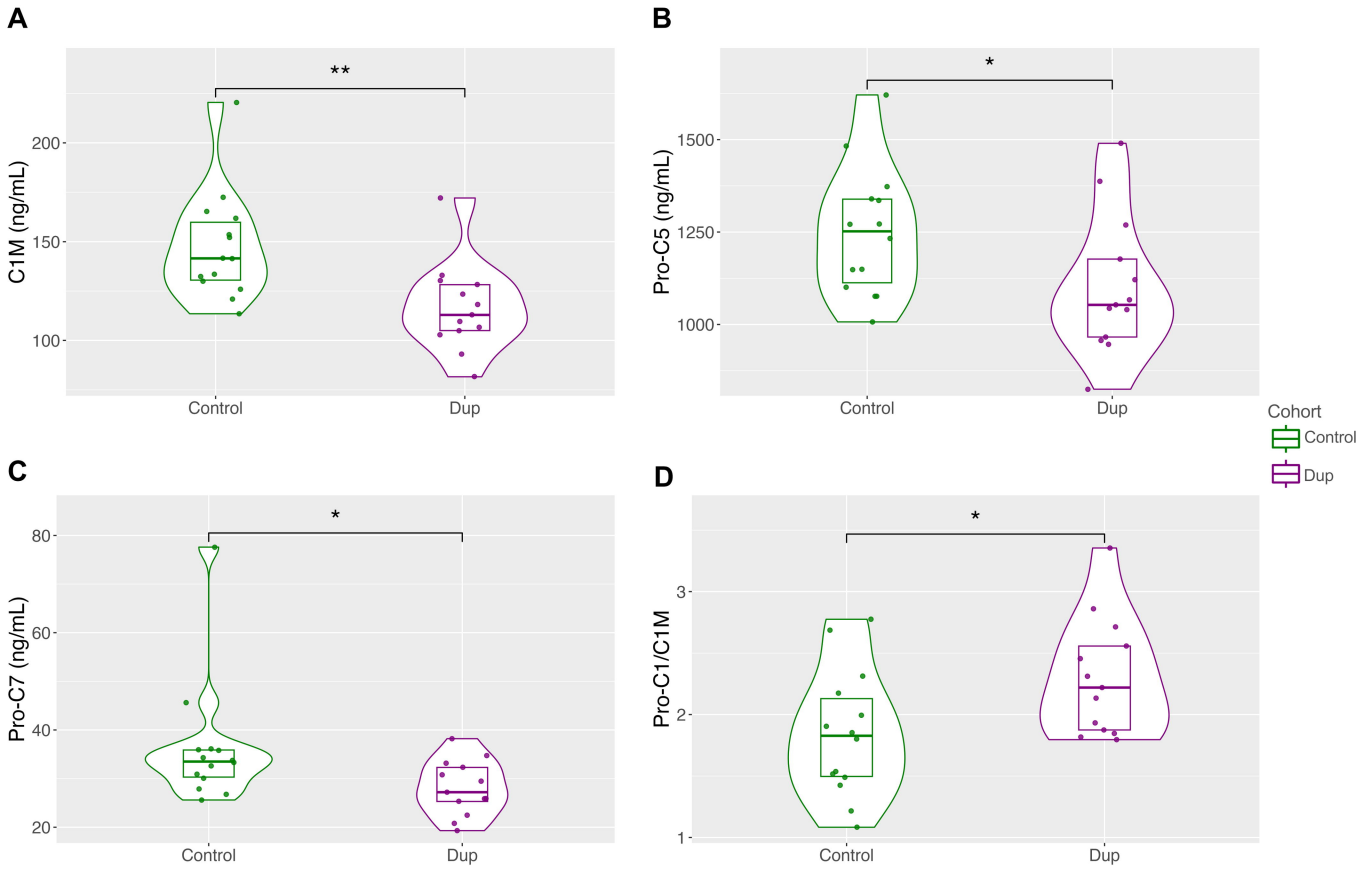
929

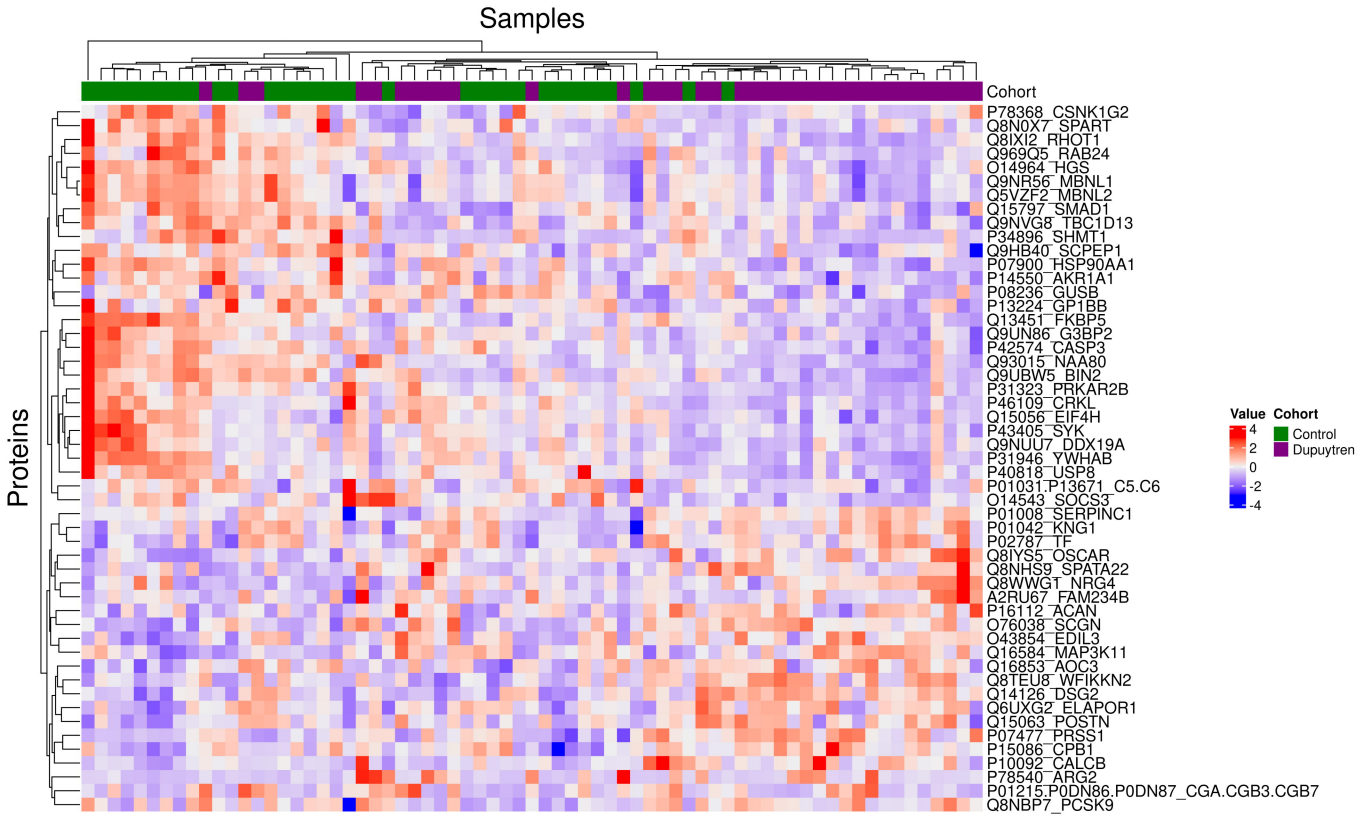
Table 4. Mass Spectrometry (MS) – SomaScan (SS) concordance

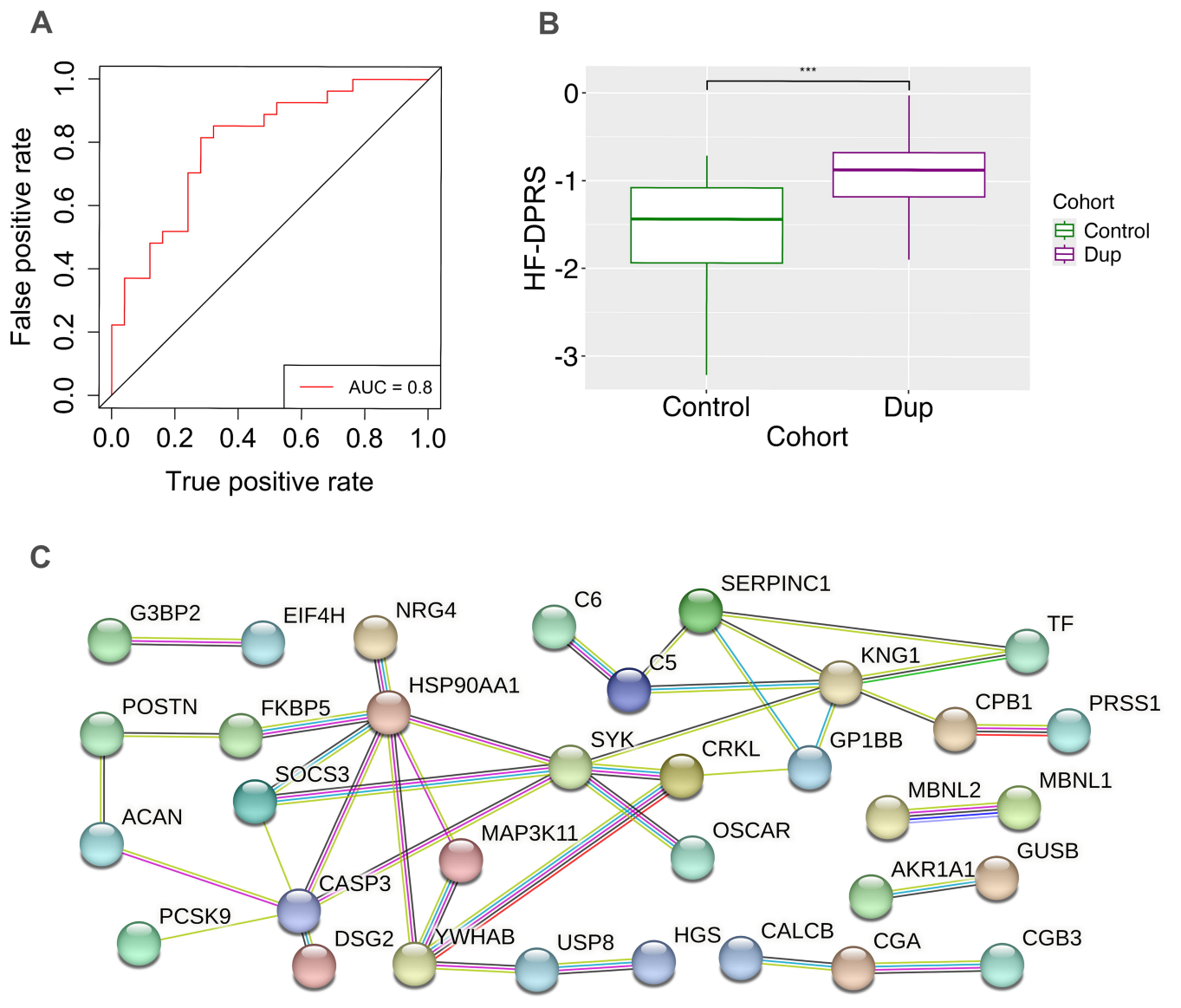
Protein	Est MS	p-val MS	FDR MS	Est SS	p-val SS	FDR SS
AOC3	0.214854	0.076413	0.232677	0.116624	0.000845	0.217972
POSTN	0.410212	0.098842	0.232677	0.182892	0.000312	0.211479
SERPINC1	0.071331	0.116338	0.232677	0.091474	0.001500	0.225251
TF	0.035200	0.567393	0.829196	0.076715	0.000025	0.172072
PCSK9	0.165811	0.755318	0.829196	0.185675	0.001371	0.222267
KNG1	0.010065	0.829196	0.829196	0.065561	0.000490	0.217972

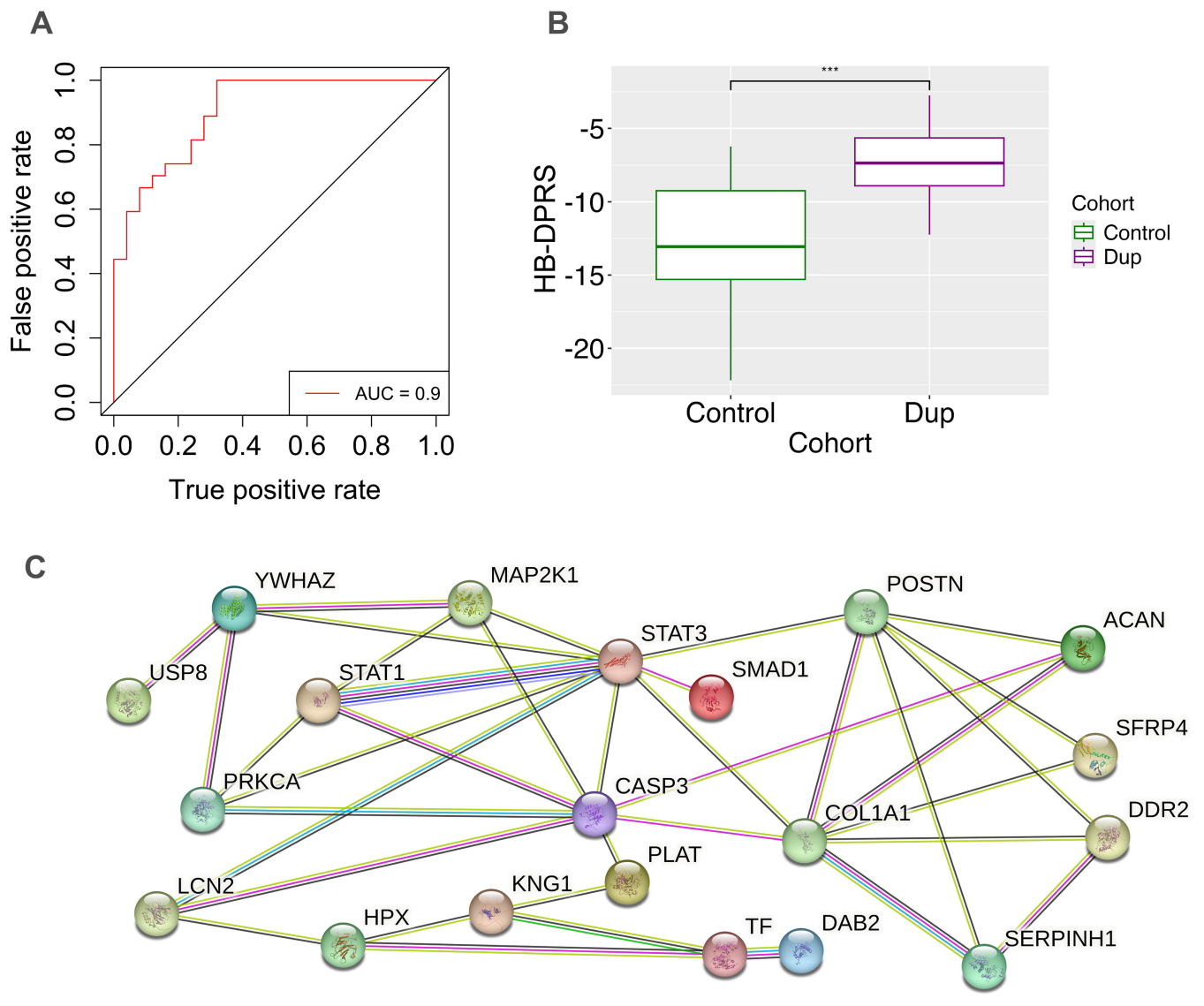
Six proteins identified by MS were also in the 54 proteins found in the hypothesis-free SS analysis. AOC3, POSTN, SERPINC1, TF, PCSK9, and KNG1. Of these, AOC3, POSTN, and SERPINC1 had False Discovery Rate (FDR, Adjusted p-value) <0.25 in both Mass Spec and SomaScan analyses.

930


A
**
C1M (ng/mL)
200
150
100
Control
Dup
B
*
Pro-C5 (ng/mL)
1500
1250
1000
Control
Dup
Cohort
Control
Dup
C
*
Pro-C7 (ng/mL)
80
60
40
20
Control
Dup
D
*
Pro-C1/C1M
3
2
1
Control
Dup


Samples
Cohort
Proteins
P78368_CSNK1G2
Q8N0X7_SPART
Q8IXI2_RHOT1
Q969Q5_RAB24
O14964_HGS
Q9NR56_MBNL1
Q5VZF2_MBNL2
Q15797_SMAD1
Q9NVG8_TBC1D13
P34896_SHMT1
Q9HB40_SCPEP1
P07900_HSP90AA1
P14550_AKR1A1
P08236_GUSB
P13224_GP1BB
Q13451_FKBP5
Q9UN86_G3BP2
P42574_CASP3
Q93015_NAA80
Q9UBW5_BIN2
P31323_PRKAR2B
P46109_CRKL
Q15056_EIF4H
P43405_SYK
Q9NUU7_DDX19A
P31946_YWHAB
P40818_USP8
P01031.P13671_C5.C6
O14543_SOCS3
P01008_SERPINC1
P01042_KNG1
P02787_TF
Q8IYS5_OSCAR
Q8NHS9_SPATA22
Q8WWG1_NRG4
A2RU67_FAM234B
P16112_ACAN
O76038_SCGN
O43854_EDIL3
Q16584_MAP3K11
Q16853_AOC3
Q8TEU8_WFIKKN2
Q14126_DSG2
Q6UXG2_ELAPOR1
Q15063_POSTN
P07477_PRSS1
P15086_CPB1
P10092_CALCB
P78540_ARG2
P01215.P0DN86.P0DN87_CGA.CGB3.CGB7
Q8NBP7_PCSK9
Value
4
2
0
-2
-4
Cohort
Control
Dupuytren


A
False positive rate
True positive rate
0.0
0.2
0.4
0.6
0.8
1.0
AUC = 0.8
B
HF-DPRS
0
-1
-2
-3
***
Control
Dup
Cohort
Cohort
Control
Dup
C
G3BP2
EIF4H
NRG4
C6
SERPINC1
TF
C5
KNG1
HSP90AA1
POSTN
FKBP5
CPB1
PRSS1
SYK
CRKL
GP1BB
SOCS3
MBNL2
MBNL1
ACAN
MAP3K11
OSCAR
GUSB
AKR1A1
CASP3
PCSK9
DSG2
YWHAB
USP8
HGS
CALCB
CGA
CGB3


A
False positive rate
True positive rate
0.0
0.2
0.4
0.6
0.8
1.0
AUC = 0.9
B
HB-DPRS
-5
-10
-15
-20
***
Control
Dup
Cohort
Cohort
Control
Dup
C
YWHAZ
MAP2K1
POSTN
ACAN
STAT3
USP8
STAT1
SMAD1
SFRP4
PRKCA
CASP3
COL1A1
DDR2
PLAT
LCN2
KNG1
HPX
TF
DAB2
SERPINH1