View Online / Journal Homepage / Table of Contents for this issue
Dynamic Article Links

Molecular BioSystems

Cite this: *Mol. BioSyst.*, 2012, **8**, 2274–2288

www.rsc.org/molecularbiosystems

PAPER

Dupuytren's disease metabolite analyses reveals alterations following initial short-term fibroblast culturing†

Samrina Rehman,[ab] Yun Xu,[d] Warwick B. Dunn,[f] Philip J. R. Day,[c] Hans V. Westerhoff,[abgh] Royston Goodacre*[bd] and Ardeshir Bayat*[e]

Received 4th May 2012, Accepted 22nd June 2012
DOI: 10.1039/c2mb25173f

Dupuytren's disease (DD) is an ill-defined fibroproliferative disorder affecting the palm of the hand, resulting in progressive and irreversible digital contracture. In view of the abnormal gene dysregulation found in DD, and its potential effect on metabolites at a functional level, we chose to examine the metabolic profile involved in DD. Using Fourier transform infrared (FT-IR) spectroscopy to generate metabolic fingerprints of cultured cells, we compared the profiles of DD cords and nodules (1) against the unaffected transverse palmar fascia (internal control), (2) against carpal ligamentous fascia (external control), and (3) against fibroblasts from fat surrounding the nodule and skin overlying the nodule (environmental control). We also determined the effects of serial passaging of the cells on DD fingerprints. Subsequently, gas chromatography-mass spectrometry (GC-MS) was employed for metabolic profiling in order to identify metabolites characteristic of the DD tissue phenotypes. We developed a robust metabolomic analysis procedure of DD using cultured fibroblasts derived from DD tissues. Our carefully controlled culture conditions, combined with assessment of metabolic phenotypes by FT-IR and GC-MS, enabled us to demonstrate metabolic differences between DD and unaffected transverse palmar fascia and between DD and healthy control tissue. In early passage (0–3) the metabolic differences were clear, but cells from subsequent passages (4–6) started to lose this distinction between diseased and non-diseased origin. The dysregulated metabolites we identified were leucine, phenylalanine, lysine, cysteine, aspartic acid, glycerol-3-phosphate and the vitamin precursor to coenzyme A. Early passage DD cells exhibit a clear metabolic profile, in which central metabolic pathways appear to be involved. Experimental conditions have been identified in which these DD data are reproducible. The experimental reproducibility will be useful in DD diagnostics and for DD systems biology.

[a] *Doctoral Training Centre ISBML, The Manchester Centre for Integrative Systems Biology, Manchester Institute of Biotechnology, University of Manchester, Manchester, M1 7DN, UK. E-mail: Samrina.Rehman@manchester.ac.uk, Hans.Westerhoff@manchester.ac.uk*
[b] *Manchester Centre for Integrative Systems Biology, Manchester Institute of Biotechnology, University of Manchester, Manchester, M1 7DN, UK. E-mail: Roy.Goodacre@manchester.ac.uk*
[c] *Quantitative Molecular Medicine Research, CIGMR, Manchester Institute of Biotechnology, University of Manchester, Manchester, M1 7DN, UK. E-mail: Philip.J.Day@manchester.ac.uk*
[d] *School of Chemistry, Manchester Institute of Biotechnology, University of Manchester, Manchester, M1 7DN, UK. E-mail: Yun.Xu-2@manchester.ac.uk*
[e] *Plastic and Reconstructive Surgery Research, Manchester Institute of Biotechnology, University of Manchester, Manchester, M1 7DN, UK. E-mail: Ardeshir.Bayat@manchester.ac.uk*
[f] *Centre for Advanced Discovery & Experimental Therapeutics (CADET), Central Manchester NHS Foundation Trust and School of Biomedicine, University of Manchester, Manchester Academic Health Sciences Centre, York Place, Oxford Road, Manchester, M13 9WL, UK. E-mail: Warwick.Dunn@manchester.ac.uk*
[g] *Netherlands Institute for Systems Biology, VU University Amsterdam, NL-1081 HV, The Netherlands*
[h] *Synthetic Systems Biology, Swammerdam Institute for Life Sciences, University of Amsterdam, The Netherlands*
† Electronic supplementary information (ESI) available. See DOI: 10.1039/c2mb25173f

Introduction

Dupuytren's disease (DD) is a benign fibroproliferative tumour of unknown aetiopathogenesis affecting the palm of the hand.[1–3] DD can cause progressive, permanent contracture of the affected digits.[2] Surgical removal of DD is often associated with a high rate of recurrence. In some cases, the severity of DD recurrence may necessitate amputation of the involved digit.[4] DD pathology is characterized by the appearance of two microscopically distinct fibrotic tissue types called the nodule and the cord.[3] Thought to be involved in the biologically most active phase of the disease, the nodule is characterised by a dense soft tissue mass containing an intense population of mostly myofibroblasts.[5] The relatively avascular and acellular cord is rich in collagen and contains fewer myofibroblasts. Although the fibroblasts are considered to be responsible for DD,[6,7] previous hypotheses have hinted at pathological relevance of adipose tissue (the fat surrounding the nodule).[8] Additionally, observations that dermofasciectomy surgery correlated with lower recurrence rates may associate the skin overlying the nodule (SON) and the fat surrounding the nodule with the development of DD.[9] Where genetic alterations

impinge on DD is unclear. In tissue biopsies and *in vitro* cultivated fibroblasts derived from DD tissue several genes were found to be over- or under expressed.[10–12] Whether, this abnormal genetic dysregulation is causal, circumstantial or a mere effect of DD and whether the observed alterations in mRNA levels had the corresponding effects at a functional level has not been fully investigated to date. This is important as the pathological formation of Dupuytren's disease and the development of contracture is directly related to the level of function associated with the aforementioned abnormal gene expresion. It is possible that causative candidates at the functional level include proteins that are responsible, or enzymes of lipid metabolism causing membrane dysfunction.

Within DD, immortalized human tissue cell lines are not yet available for the purpose of scientific research. Thus, primary cultures are employed and the DD cells may lose their unique disease characteristics when cultured, or develop artifact differences when compared to normal cells. These differences may have little to do with the difference between the diseased and the healthy tissue in the patient, and may rather represent positive selection for cells able to survive successive *in vitro* cell culture passages. In view of the above, we hypothesize: (i) that the culturing conditions have a profound impact on gene expression and metabolic activities; yet (ii) for a limited period of time, cells taken into culture retain DD characteristics. If proven correct, this should lead the way to a robust experimental system for studying DD in more depth that will lead to direct assessment of DD and unique disease characteristics.

Systems biology (SB) enables interpretation of measurements taken with different analytical tools of different functional levels which relate transcriptional changes to changes at these levels.[13,14] Within the context of assessing metabolism many strategies are used.[15,16] With reference to the current study, these include platforms such as Fourier transform-infrared (FT-IR) spectroscopy (to acquire metabolic fingerprints (global, high-throughput, rapid analysis of whole cells to provide sample classification used as a screening tool to discriminate between samples from different biological status) and non-invasive metabolic footprint screening (analyses of the (exo) metabolites secreted/excreted by the cell)[17] as potential diagnostic and screening tools, and gas chromatography-mass spectrometry (GC-MS) for quantification and identification of metabolites.[18,19] In its analysis of how the nonlinear networking of molecules produces *in vivo* function, systems biology is sensitive to differences between the *in vivo* physiological state and experimental *in vitro* states. Metabolites are a dynamic and sensitive measure of the phenotype of living systems that results from the interactions between genes and environment managed by the system itself.[16] Accordingly measurement and analysis of metabolites can be a precise and potentially valuable resource for (i) evaluation of DD phenotypic changes when DD cells are grown in cell culture, (ii) identifying biomarkers for diagnostics[20] and (iii) for identifying and linking genetic changes underlying malfunction with metabolites.[21] An increasing number of success stories demonstrate metabolomics is contributing to the understanding of a number disease processes.[22–24] There have been no previous reports however, of metabolomic studies of DD, neither of the metabolic fingerprint (intracellular) nor of the metabolic footprint (extracellular) type associated with DD.

In this study, we therefore embark on a systems analysis of DD that starts at the metabolic level. We examined whether DD cells in culture differed metabolically from non-DD cells and measured this difference if persistent during serial passaging. We carried this out by implementing (i) the concept of a systems signature defined through FT-IR spectroscopy and (ii) metabolic profiling achieved with GC-MS.

Method

Experiment 1: whole-cell fingerprinting using FT-IR spectroscopy

Patient recruitment. Metabolic fingerprint and reproducibility studies were investigated with samples from patients in cohorts A and B. The metabolic footprint was investigated with culture media from samples in cohort A. Metabolic profiling was performed on samples from patients in cohort C. The age and demographics for all patients can be seen in supplementary Table S1.

Study 1, 3-5, cohort A. All cases involved in the study were diagnosed to have advanced stage DD, which was determined by the presence of nodule and cord causing contracture of the metacarpophalangeal joint and the proximal interphalangeal joint in the involved hand. The mean age of the patients participated in Study 1 was 67 ± 10 years. All patients were male Caucasians who had not undergone any previous surgical or non-surgical treatments with exception of patient DD13 who was female.

Study 1-4, cohort B. Three DD cases and three controls subjects (CTD) were included in the study. All recruited DD cases were diagnosed with advanced stage of DD (Grade 3), which was determined clinically by an experienced hand surgeon. All patients presented flexion contracture of the metacarpophalangeal joint and proximal interphalangeal joint as well as the presence of nodules. All DD patients in this study were male Caucasians and one (DD2) had undergone previous surgical treatment. The mean age was 69 ± 2 years. Two of the control subjects were male and two female. The average age of the control subjects was 55 ± 21 years. The study was approved by the local and regional ethical committee for human subjects' research.

Sample collections for reproducibility studies. DD tissue phenotypes (nodule, cord, unaffected transverse palmar fascia, subcutaneous fat superficial to nodule and SON) were carefully dissected using magnifying loupes from each patient at the time of surgery (supplementary Fig. S1A). Tissue biopsies from carpal ligamentous fascia, were obtained from individuals undergoing carpal tunnel release (supplementary Fig. S1B). Each biopsy was bisected for cell culture processing and histology tissue processing. The biopsies used for establishing tissue cultures were thoroughly washed for 15 min in 1 × Dulbecco's phosphate buffered saline (Lonza, Belgium) and 1% penicillin/streptomycin (Lonza, Belgium), at room temperature. For histological analyses to determine tissue

and cell morphology, the biopsies were stored in formalin at 3 °C, and processed within 48 h.

Specimen processing & tissue/cell culture for whole cell fingerprinting using FT-IR spectroscopy. To establish tissue culture, the biopsies were further dissected into small pieces, roughly 1 mm^3 in size, with sterile scalpels. The tissue pieces were incubated in 0.25–5% collagenase A solution (Roche Diagnostics, GmbH, Germany) at 37 °C for 2.5 to 3 h. The collagenase activity was inhibited using fibroblast-culturing media (Dulbecco's Modified Eagle's Medium 3 (Lonza, Belgium) supplemented with 10% heat-inactivated fetal bovine serum (Lonza, Belgium), 1% penicillin/streptomycin (Lonza, Belgium) and 1% non-essential amino acids (Lonza, Belgium)). The digested samples were centrifuged at 1500 rpm (*ca.* 400 × g) for 5 min. Each pellet was re-suspended in 5 mL fibroblast culturing media, seeded to 25 cm^2 culturing flask (Corning, UK) and incubated at 37 °C in 5% CO_2. The culturing media was replaced every 48 h and cell passages were carried out at approximately 80–90% confluency using trypsin-ethylene diamine tetraacetic acid (200 mg L^{-1} ethylene diamine tetra-acetic acid, 500 mg L^{-1} trypsin; Lonza, Belgium). The first sub-culturing (passage) was performed on cultures that were grown directly from the biopsies. These cultures were called passage 0 (P0). Following the first centrifugation of cells, $\frac{1}{2}$ of the pellet obtained was seeded onto 1 × 75 cm^2 culturing flask in Study 2 (now called P1 cells) and the remaining $\frac{1}{2}$ pellet was frozen in DMSO containing freezing media. (In Study 1, all cells from the pellet were passaged onto new flasks for the first time (4 × 25 cm^2 in equal amounts) and no sample from the previous passage (P0) was retained. Further passaging was performed on $\frac{1}{4}$ of the cells grown from the current passage (called P1) and $\frac{3}{4}$ was kept in freezing media containing DMSO into 3 separate Nunc tubes equally and transferred into Mr Frostie at room temperature. These were then stored at −80 °C until required for FT-IR analysis. During the incubation period the fibroblasts purity was assessed by morphological observation under an inverted phase contrast microscope. The spent culture medium containing all excreted metabolites (footprint) were kept in 15 mL falcon tubes and stored at −80 °C until analysis. The approximate percentage confluence (85–90%) and number of days that each sample was passaged was recorded. All passages of the cell cultures were used in this study. All work was conducted using a single production batch of serum.

Morphological assessment and haemocytometer counting. Morphological changes of DD and CTD fibroblasts in all groups were monitored under an inverted phase contrast microscope during the tissue culture experiments. At the end of each passage, cells were washed with phosphate buffered saline (PBS), and detached from the culture dish by 0.25% trypsin. Cell number was then counted through direct visualisation using a haemocytometer. Prior to sub-culturing, 20 μL of cell suspension was mixed with 20 μL Trypan Blue (Nacalai Tesque, Inc.), a dye exclusion that leaks into cells with damaged plasma membranes. In this way, the dead cells were stained blue, allowing the living and the dead cells to be distinguished. A 10 μL amount of this solution was then placed in a haemocytometer and the number of living cells counted, while being viewed with a light microscope. Average number of cells per T25 cm^2 flask were 600–650 000 for nodules, 400–500 000 for cords, 350–400 000 for the internal control fascia, 400–500 000 for the fibroblasts derived from the subcutaneous fat tissue and 600–670 000 for the skin overlying the nodules. The average number of cells per flask and the number of days between each passage for each sample was recorded.

FT-IR spectroscopy sample preparation. 15 μL aliquots of the sample cells dissolved in 50 μL DPBS suspensions were evenly applied onto a silicon (Bruker Spectrospin Ltd., Coventry, United Kingdom) microplate containing 96 wells. Each sample was spotted in triplicates in a random arrangement. Prior to analysis the samples were oven dried at 50 °C for 30 min. The FT-IR instrument used was an Equinox 55 infrared spectrometer (Bruker Spectrospin Ltd.) equipped with a high-throughput motorized microplate module, HTS-XT under the control of a computer program with OPUS software version 4. A deuterated triglycine sulfate (DTGS) detector was employed for transmission measurements of the samples to be acquired. Spectra were collected over the wavenumber range of 4000 cm^{-1} to 600 cm^{-1}. Spectra were acquired at a resolution of 4 cm^{-1}. To improve the signal-to-noise ratio, the spectra were co-added and averaged. Cohort A; 468 spectra (78 samples spotted in triplicate and spectra acquired twice) and cohort B; 927 spectra (103 samples spotted in triplicate and spectra acquired thrice). Each sample was thus represented by a spectrum containing 1764 points and spectra were displayed in terms of absorbance calculated from the absorbance spectra using the Opus software. The combination of both biological and analytical replicates was employed in the FT-IR analysis to measure the biological variance in the data. The employment of 'analytical or machine replicates' was averaged to reduce the heterogeneity of the technical replicates in the analysis of the data. Prior to spectral collection an empty well was used to take a reference background measurement.

FT-IR-footprint sample preparation. 20 μL of the secreted metabolites collected in spent media were spotted directly onto the silicon microplates, oven dried and subjected to FT-IR as above.

FT-IR Data pre-processing and multivariate statistical analysis. All statistical analyses were performed using Matlab R2010a (MathWorks, Inc., MA). The ASCII data were imported into Matlab. To minimise problems arising from baseline shifts, empirical pre-processing techniques to reduce/eliminate light scattering effects were applied. The following procedures in Matlab were implemented. To reduce/eliminate light scattering effects extended multiplicative scatter correction (EMSC),[25] was applied to the spectra. These EMSC corrected spectra were then detrended by subtracting a linearly increasing baseline from 4000 to 600 cm^{-1} where necessary. Data were then grouped into categories listed in various combinations as shown in supplementary Table S2 for comparisons using multivariate analyses (MVA). The pipeline used for MVA on the hyperspectral FT-IR data is illustrated in supplementary Fig. S3.

To reduce the dimensionality of the multivariate data whilst preserving most of the variance, Matlab (and some figures were also created in PyChem (3.05a))[26] was used to perform principal component analysis (PCA).[27,28] For each PCA performed on the chosen combination of samples, of the original 1764 spectral points in the respective study, the first 20 principal components (PCs) were recorded for subsequent analysis by discriminant function analysis (DFA). This was employed for samples where PCA was insufficient to show clear pattern recognition from the scores plot. DFA (also known as canonical variates analysis[29]) then discriminated between groups on the basis of the retained PCs that were used as inputs to the DFA algorithm with the *a priori* knowledge of which spectra were replicates. DFA was programmed to minimise 'within-group' variance and maximise 'between-group' variance. This process uses information based on the biological replicates from each sample, rather than its class membership (that is to say, the algorithm was not given any information on the passage number nor sample type (nodule, cord etc)) it therefore acts as an efficient tool to minimise the experiment variations and normally does not bias the analysis, allowing any natural trends or time-dependent trajectories to be observed. In PyChem,[26] the Euclidean distance between *a priori* group centres in DFA space using the first two functions (DF1 and DF2), was used to construct a similarity measure, and these distance measures were then processed by an agglomerative clustering algorithm to construct a dendrogram (not shown).

Experiment 2: Metabolic profiling of fibroblasts using GC-MS

Patient recruitment

Study 6, cohort C. DD patients ($n = 3$) were used in this study. Patient recruitment procedure was followed as previously stated and demographic information including age is given in supplementary Table S1. supplementary Fig. S4 shows the flowchart of steps involved in this experiment, from sample collection to data analysis.

Source of biopsy tissue specimens. Tissue biopsies were obtained from 3 male DD patients having dermofasciectomy: nodules from the palm ($n = 3$), cords from the palm ($n = 3$) and transverse palmar fascia (internal control ($n = 3$). Biopsies were harvested at the time of surgery. Tissue was carefully excised to include the diseased or the normal fascia without the adjacent adipose/connective tissue. The tissue was then placed immediately into DPBS and transported to the laboratory within an hour to establish cell cultures.

Cell culture. A total of 27 samples were processed and stored. Samples from (passage 1) DD nodules ($n = 3$), cords ($n = 3$) and fascia ($n = 3$) were seeded into T25 cm^2 culture flasks. Upon 90% confluence, samples were passaged into $2 \times$ T75 cm^2 culture flasks. Approx 1–1.5 million cells were obtained from each T75 cm^2 flask. Upon 85–90% confluent, these were sub-cultured into 3×150 cm^2 flasks; 3 replicates for each sample. DD fibroblasts were grown until 85–90% confluent in 21% O_2 and 5% CO_2. Culture medium for all conditions had the same formulation; *i.e.*, DMEM (500 mL) supplemented with L-glutamine (1%), NEAA (1%) and FBS Gold (10%) and penicillin/streptomycin (1%). After 48–72 h, culture medium was replaced with fresh medium. Trypsinisation of adherent cells was avoided during passage and large cell scrapers were used to harvest cells, which were seeded into new flasks. All work reported was conducted using a single production batch of media components.

Metabolite quenching & extraction. Cells were grown from each biological sample until 85–90% confluent. The media was aspirated and the cells were washed once with 4 °C PBS (12 mL) and quickly aspirated. 70% methanol in water (8 mL pre-chilled in dry ice) was added immediately to quench metabolic activity. A large cell scraper was used to harvest cells quickly over ice. The cellular biomass was removed by pipette aspiration and collected in 15 mL centrifuge tubes. Metabolites were immediately snap frozen in liquid nitrogen. Metabolites were then extracted through 3 freeze-thaw cycles (vortexed for 30–40 s each time and thawed on dry ice) in order to permeabilise the cells, resulting in the release of the metabolites from the cells. The supernatant (containing the metabolite complement) was collected by centrifugation ($15\,000 \times g$ for 7 min) and transferred in to pre-weighed falcon tubes and placed in dry ice. The extracts were stored at −80 °C for 48 h.

Derivatisation of metabolites for GC-TOF-MS analysis. Samples were prepared for MS by determining the accurate volume of extracts to dry down. Three replicate pellets for each sample containing cell debris were dried to remove residual solution and then weighed to determine the mass of dried biomass after extractions to determine a normalised volume of metabolite extracts required for analysis for each sample. A 0.3 mg ml^{-1} solution of succinic d_4 acid in water was added to each solution as an internal standard. These volumes were then lyophilized for 16 h in a vacuum concentrator (HETO VR MAXI with RVT 4104 refrigerated vapor trap; Thermo Life Sciences, Basingstoke, U.K). For MS analysis dried extracts were then derivatized as follows; 50 μL of 20 mg mL^{-1} *O*-methoxylamine hydrochloride in pyridine was added, vortexed, and incubated at 60 °C for 30 min in a dri-block heater. A volume of 50 μL of MSTFA was then added and the extracts incubated at 60 °C for a further 30 min. On completion, 20 μL of retention index marker solution was added (0.3 mg mL^{-1} docosane, nonadecane, decane, dodecane, and pentadecane in pyridine) prior to centrifugation at $15\,800 \times g$ for 15 min. The resulting supernatant (90 μL) was transferred to GC-MS vials for analysis.

Gas chromatography-time-of-flight mass spectrometry analysis. The samples were analyzed in a random order by employing a GC-TOF-MS (Agilent 6890 GC coupled to a LECO Pegasus III TOF mass spectrometer) using a previously described method.[30,31]

Metabolite identification. The GC-MS data were deconvoluted producing a peak table of the metabolites identified, a three dimensional matrix of information: retention time (related to the time since injection), m/z ratio and chromatographic peak area (defined applying a single quantification m/z). Raw data were processed using LECO ChromaTof v2.12

and its associated chromatographic de-convolution algorithm, with the baseline set at 1.0, data point averaging of 3 and average peak width of 2.5. A reference database was prepared, incorporating the mass spectrum and retention index of all metabolite peaks detected in a random selection of samples so to allow detection of all metabolites present. Each metabolite peak in the reference database was searched for in each sample and if matched (retention index deviation $< \pm 10$; mass spectral match > 750) the peak area was reported and the response ratio relative to the internal standard (peak area-metabolite/peak area-succinic-d_4 acid internal standard) calculated. These data (matrix of N samples $\times$ P metabolite peaks) representing normalised peak lists were exported in ASCII format for further analysis. Metabolites were definitively identified following MSI[32] by matching the mass spectrum and retention index of detected peaks to those present in a mass spectral library constructed at the University of Manchester.[33] A match is defined as a match factor greater than 750 and a retention index ± 10.

Chemometric analyses of GC-MS raw data and metabolite levels. Univariate and multivariate analyses were performed on the response ratio data sets and also followed MSI reporting standards.[34] All statistical analyses were performed using Matlab R2010a (MathWorks, Inc., MA). Within a GC-MS-based data matrix composed of response ratios (peak area-metabolite/peak area – internal standard), it is possible to obtain zero (or not detected) values for any given metabolite peak caused by either of the following reasons: the metabolite is not present or is present at a concentration below the limit of detection; or the metabolite cannot be resolved from others in the chromatograph by the deconvolution software. In these cases, the following procedure was used to improve data structure for statistical analysis techniques. For univariate analysis, two approaches were applied: (1) All data was accounted and the zero values were replaced with 'NaN' (not a number); (2) only the median value of the three biological replicates was accounted. All peaks with more than 20% missing values were removed from the analysis. Outliers were suppressed using a trimming procedure by excluding extreme data points from raw PCA scores plots. For multivariate analysis, if two of three replicates were zero values and the third replicate was a non-zero value, the third (non-zero) replicate was replaced with zero. If two of three replicates produced values and the third replicate was a zero value, the zero value for the third replicate was replaced with the mean of the other two replicates. This is to reduce the influence of non-detected metabolites due to analytical reasons (*e.g.* co-eluting, low concentration *etc.*).

Prior to multivariate statistical analyses, data were normalised to zero mean unit variance, so that results were not dominated by a small number of high intensity peaks but gave equal weighting to peaks of low intensity. MVA were then performed using PCA and its modified algorithm which aimed to cope with the study with multiple influential factors named Analysis of variance – PCA (ANOVA-PCA) to quantify the relative variance arising from the three sample types (*i.e.* nodule, cord and transverse palmar fascia), patients and analytical uncertainty and to test the significance of differences in the chemical composition.

Univariate statistical analysis was also performed using the non-parametric Wilcoxon sign rank test (Table 1) and Friedman test to identify metabolites, which showed significant difference between two types of samples. The critical p-value for rejecting the null hypothesis in a single test was 0.05. Boxplots were drawn to show relative concentration distributions for all samples with respect to a given peak.

Results

Histology

The DD nodule, cord, fat, the skin overlying the nodule, and the internal (the transverse palmar fascia) and external (the carpal ligamentous fascia) control from healthy individuals were identified and distinguished. The histology slides were confirmed by an independent experienced histopathologist. The nodules displayed regions of high cellularity and lots of nuclei were stained/visible, while the cords displayed a tendon like arrangement. Cross sections from the fat and the skin overlying the nodule were also examined supplementary Fig. S2.

Whole-cell FT-IR fingerprinting and footprinting of cell cultures

FT-IR spectral collection. Cell samples that had been washed thoroughly to remove any extracellular metabolites and dried, were investigated by FT-IR spectroscopy. The raw and normalised absorbance FT-IR spectra for DD and control fibroblasts are shown in Fig. 1A–C. These spectra display the typical fingerprint of each sample and the footprint in the case of spent media. These (and indeed all FT-IR spectra) show broad and complex contours with little qualitative difference between the spectra. The spectra contain information on functional group vibrations resulting in the absorbance of infrared light at specific wavenumbers ($1/\lambda$).[17] Some prominent regions (Fig. 1C) arose from vibrational modes of protein (O–H stretch centered at 3400 cm^{-1}), fatty acids (CH_x stretches at 2956–2850 cm^{-1}), and proteins (amide I, C=O at 1652–1648 cm^{-1}; amide II N–H, C–N at 1550–1548 cm^{-1}). A mixed region from 1460–110 cm^{-1} contains information from fatty acids, polysaccharides, nucleic acids, proteins and polysaccharide rings and further C–O vibrations at 1085–1052 cm^{-1}.[17] Both large molecules such as proteins and nucleic acids as well as small constituents were detected because the sample preparation used simply DPBS to suspend the cells and the samples. FT-IR spectra for various DD phenotype fibroblasts (from all five anatomical locations), fibroblast growth media, and freezing media were also investigated. Again, all spectra showed broad and complex contours, in which there was little qualitative difference between the spectra.

Multivariate statistical analyses to determine variability of samples. A total of 234 samples from cohort A were each analysed twice. The 309 samples from cohort B were each analysed three times. PCA was performed in various combinations on the data set, across samples selected on the basis of passage numbers, patients and anatomical sites.

The number of significant principal components (PCs) were determined in three ways. Firstly, the scree plot of eigenvalues,

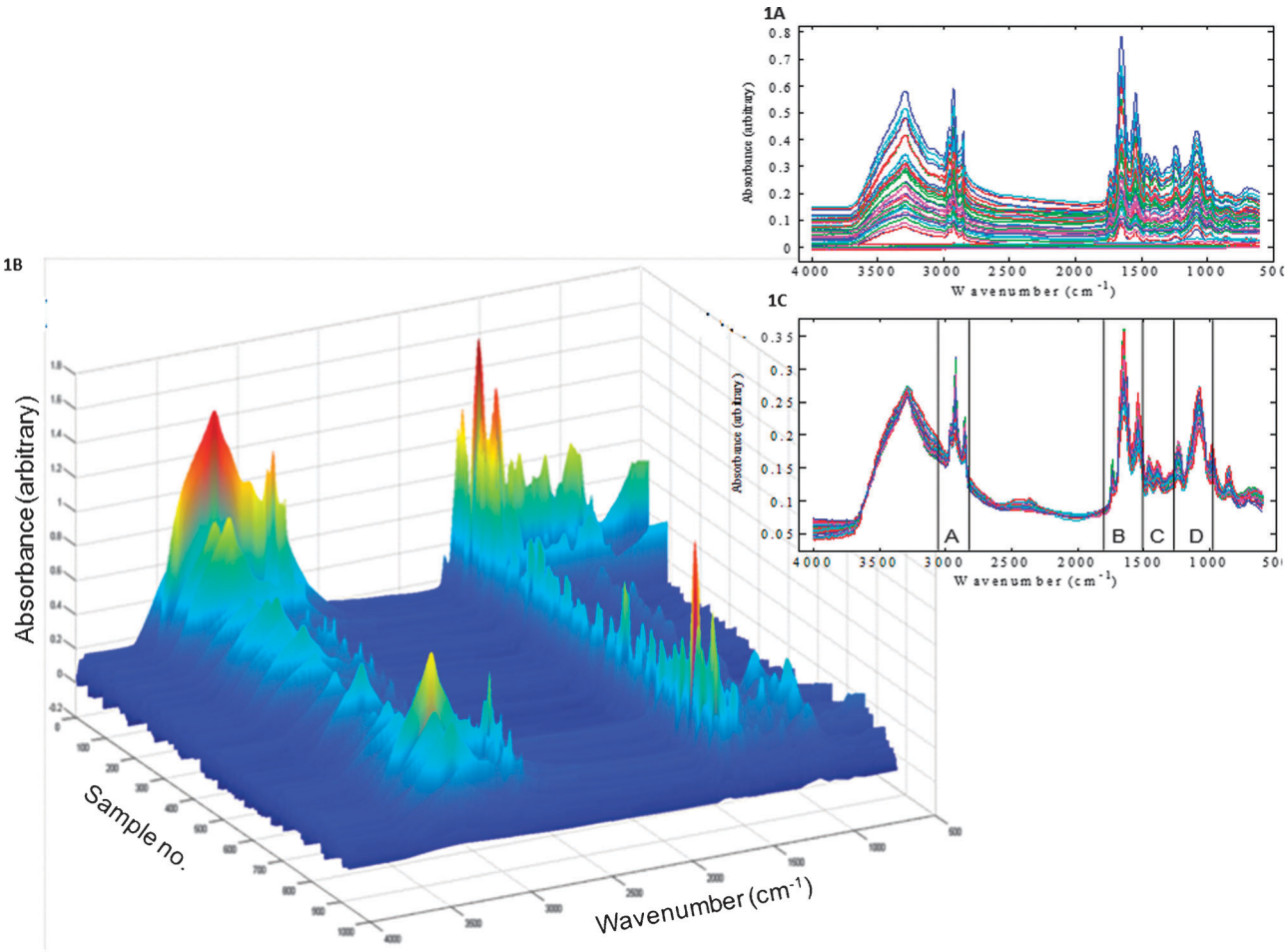

Fig. 1 A and B The raw Fourier transform infrared (FT-IR) spectra from metabolic fingerprints of cultured cells. Each sample is represented by an absorbance *vs.* wavenumber spectrum. B illustrates a 3-D raw FT-IR spectra with 927 samples from Study 2. C illustrates a processed FT-IR spectra of cultured cells with band assignments. Samples were normalised using the extended multiplicative scatter correction preprocessing method and baseline correction. These spectra have been offset to see the features more readily. Key to vibrational bands: A = fatty acid, B = amide, C = mixed, D = polysaccharide.

was examined. Its shape suggested selecting between two and three PCs; the majority of the variance within the data set was captured by the first few PCs with little importance on the later PCs. Secondly, the score plots for PC1 & PC2, PC1 & PC3, PC2 & PC3, and subsequent PCs were analysed. Third, the percentage cumulative variance plot was examined. It was found that for PCs $>$ 3 the plots could not be distinguished from noise. On the basis of these results it was decided that 3 PCs would be the best choice. This selection means that most of the variance in absorption of the examined spectra will be further interpreted in terms of the above three factors. These will then be related to: (i) passage number *vs.* sample type; (ii) individual patients *vs.* sample type including all passages. The FT-IR experimental sample sets were divided into studies (Experiment 1.1–1.5, Supplementary Table S2). Fig. 2, 4, 5 represent the PCA scores plots from PC1 and PC2 from experiments 1.1, 1.3 and 1.4. (supplementary Table S2), and Fig. 3 represents the PC-DFA scores plot for the samples from primary cultures from experiment 1.2 (cohort B). They show the relationships between the three different phenotypes and confirm their metabolic differences. Each numeric code represents a single biological sample. The closer the samples cluster together the more biochemical similarity they possess.

DD nodule and DD cord *versus* transverse palmar fascia (Experiment 1.1). PCA on normalised data was performed to test whether the quality of the spectral data was good enough to determine covariance of fibroblast samples that lead to clustering. Fig. 2A shows the scores plots for DD and healthy in samples from cohort A.

There was a clear separation of control samples from disease samples. In addition DD nodule and cords show an additional level of heterogeneity. There were two distinct clusters identified among the DD samples separating them into cord and nodule (circled in green and red) groups, both from the transverse palmar fascia (circled in blue) in passage 1.

In cohort B analysis, cultures from passage 0 were included (Fig. 2B). Although very little biomass was harvested (for this 'passage 0' $\frac{3}{4}$ of the initial monolayer grown from primary fibroblast was subjected to FT-IR spectroscopy), a clear separation between control and DD disease samples can be observed for these passage 0 profiles. This continues to be so

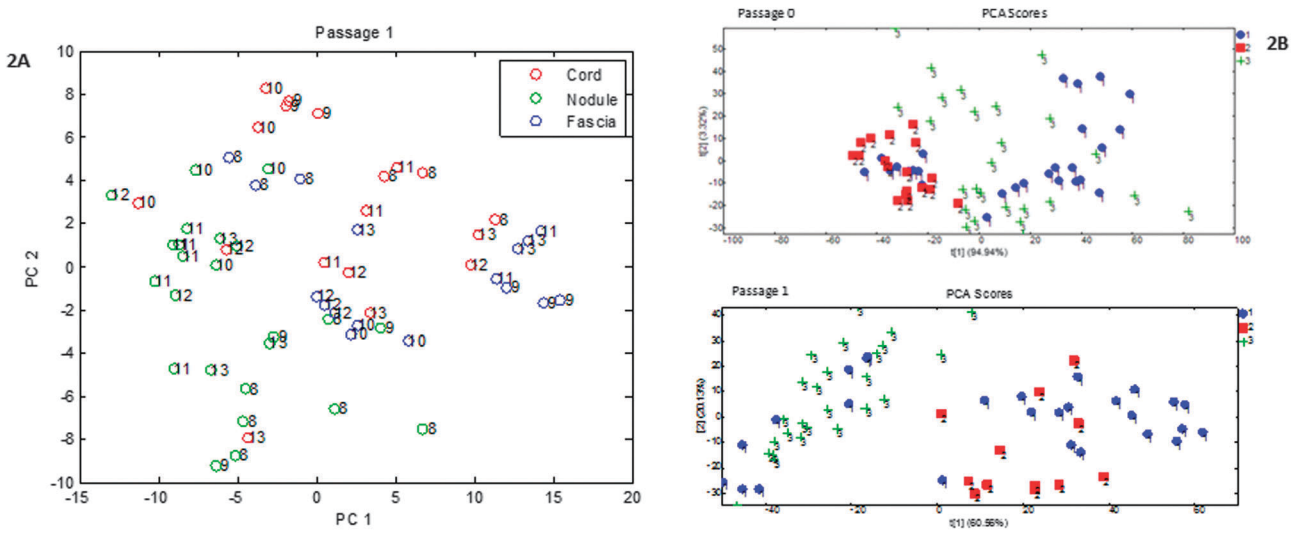

Fig. 2 A This figure illustrates projection of the fourier transform infrared (FT-IR) spectra of Dupuytren's disease (DD) fibroblasts derived from the nodule, cord and fascia (internal control) onto the plane defined by PC1 and PC2. Patients are labeled by their numbers *e.g.* for each patient where the three cell types were analysed (*e.g.* DD nodule, cord and fascia), three different coloured circles represent each tissue type. Red circle = cords, green circle = nodules, blue circle = fascia (internal control). B This figure illustrates projection of the FT-IR spectra of DD fibroblasts derived from the nodule, cord and fascia (internal control) from passage 0 and passage 1 onto the plane defined by t[1] (PC1) and t[2] (PC2). The samples are labeled by numbers and shapes; 1 blue circle = transverse palmar fascia, 2 red square = nodule, 3 green cross = cord.

for the passage 1 sample, with some clusters of fascia entering the nodule space in PC1.

DD nodule and DD cord *versus* internal (transverse palmar fascia) and external (unaffected normal palmar fascia) control (Experiment 1.2). Fig. 3 shows the DF1 *vs.* DF2 score plot from the DD subsets compared to internal and external control fascia. Its analysis along the DF1 axis shows that this component divides the whole set of spectra into four groups; a cluster of red circles (DD nodules), a majority of blue circles, (DD fascia-internal control) on the right half of the plot, a majority of turquoise/cyan circles (CTD fascia-external control) on the left half of the plot, and green circles (DD cords) closer to the centre (zero on *x*-axis): There was clear separation between the DD fascia, DD nodule, DD cord, and also the external CTD control. Clusters of samples from DD cells always clustered apart (while showing separation between the individual samples) from those of the CTD. This separation suggests that internal fascia is an appropriate control and can be distinguished from diseased fibroblasts using chemometrics

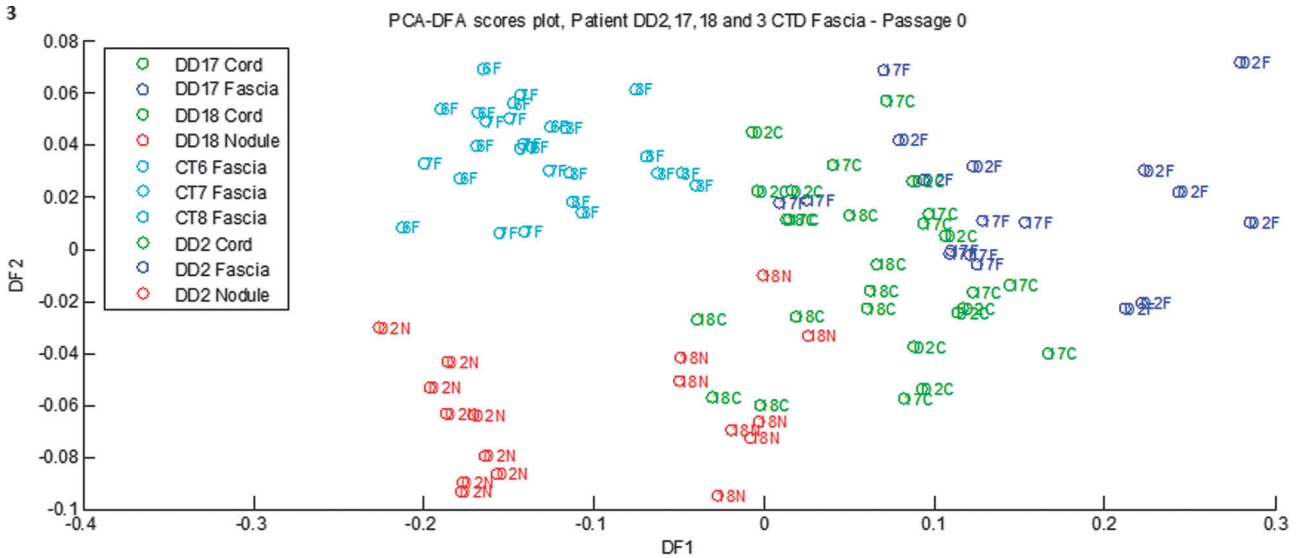

Fig. 3 Projection of the Fourier transform infrared FT-IR spectra of Dupuytren's disease (DD) fibroblasts derived from the nodule, cord, fascia (internal control) and carpal tunnel disease (CTD) palmar fascia (external control) from passage 0 onto the plane defined by DF1 and DF2. Patients are labeled by their numbers *e.g.* for each patient where the four cell types were analysed (*e.g.* DD nodule, cord, fascia and CT external fascia), there are four different colours to represent them; blue circle = fascia (internal control); red circle = nodules; green circle = cords; and turquoise circles = CTD fascia (external control).

procedure put together here. The use of internal fascia as the control should contribute to homogeneity in future studies.

DD nodule and DD cord *versus* transverse palmar fascia, fat, skin overlying nodule (Experiment 1.3). PCA scores of the five different anatomical locations were examined. Clear separation of DD nodule, DD cord and the transverse palmar (non-DD) fascia were observed from fibroblasts derived from fat and SON cell types in all patients. Fig. 4A and B demonstrate this for samples from a single patient across PC2. Fascia clusters can also be separated from nodules but are closer to the cords. Separation of fat from SON is observed in PC3. The PCA scores plot did not show clear separation between clusters from sample sets combined from multiple patients. This shows that the high biodiversity between patients also affects the metabolic variables correlating with DD, and again highlights the importance of internal controls. In some cases, the nodule and SON derived fibroblasts were in closer proximity than those from cultures derived from cords, fascia and fat.

Passage effect on metabolic fingerprints (Experiment 1.4). As above, PCA was performed to test the quality of the FT-IR spectra (Fig. 5). As the number of passages increased the samples became increasingly random in terms of the PCA results and the DD samples overlapped increasingly with those derived from other tissue phenotypes. In cohort A samples, the separation between the clusters of individual phenotypes (nodule, cord, fascia) became very small for samples of passage number = 2. Samples from cultures of passage numbers $\geq$ 3 did not demonstrate robust separation. Although some separation was observed, these late-passage samples were not super-imposable on cultures derived from the same phenotype in passage 0–2, and thereby lost distinction between disease and control phenotypes. PCA score plots from passages 5 and 6 are shown in Fig. 5A. The change with cell passaging can also be observed in a single patient: Fig. 5B demonstrates this trend for the three sites. For passage number 1 clusters from nodule, cord and fascia are distinct, but as passage number increases, the clusters of the different classes move closer together and within classes are no longer clustered as tightly.

As there is biological variance between patients, this difference is not as clear in PCA scores plots alone, therefore PC-DFA was performed. However, on inspection of the PC-DFA scores plots (data not shown) the results from Cohort A implied that performing serial passages $>$ 3 did not produce significant separation between the metabolic profiles obtained from spectral data of the three different sites (nodule, cord and fascia).

In cohort B a rather inconsistent sample size was analysed. This inconsistency was due to several factors ranging from a smaller number of DD biopsies, cells that did not grow or survive the tissue processing step, samples arriving on separate occasions with up to 2–4 weeks gap. Still, in early passage numbers clear separation of control samples and DD disease tissue was observed. In addition, DD fibroblasts (nodule and cords) showed an additional level of heterogeneity: There were 3 distinct clusters between passage 0, 1, 2 samples and again randomisation started to occur in samples from cultures from passage 3 (Fig. 5C).

Finally, Fisher's ratio was calculated for each patient to determine separation/variability within class. The larger this ratio is, the better is the separation between classes (*i.e.* well separated classes and tight clusters for each class). The relation between Fisher's ratio and passage number is plotted in Fig. 6 for patients in cohort A. This ratio decreased linearly with

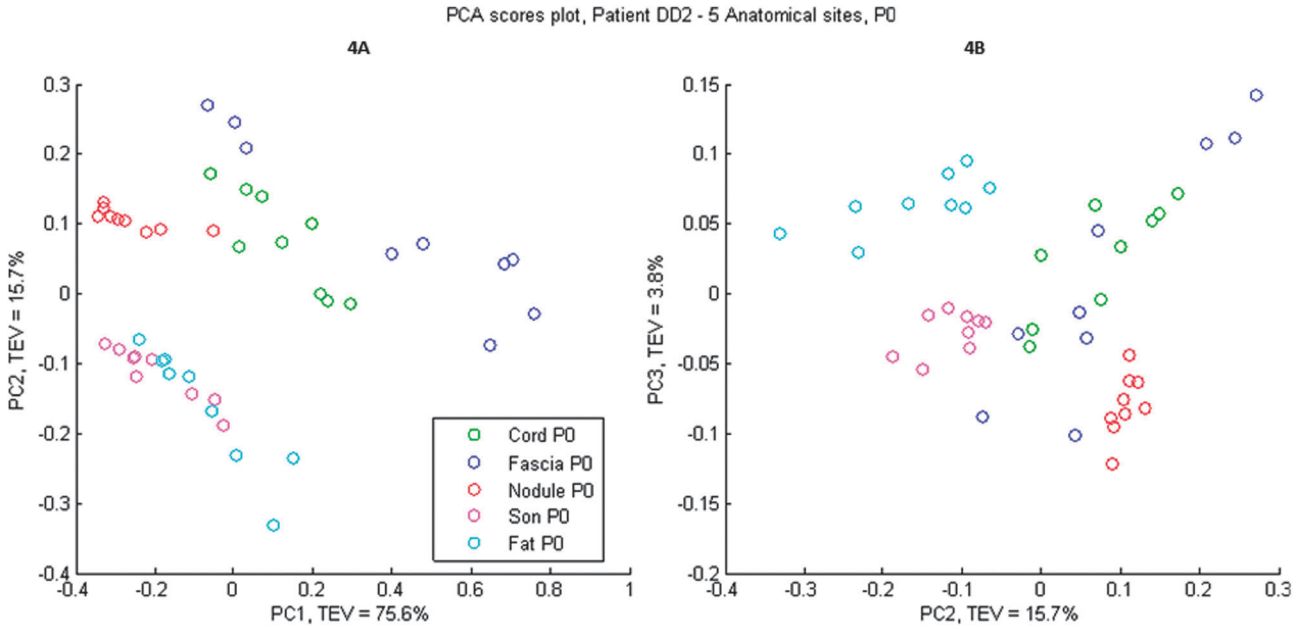

Fig. 4 A PCA scores plot of the five different sites from a single patient; Dupuytren's disease (DD) DD2, distinct clusters are observed, the spectral profile of skin over nodule (SON) and fat are highly similar to each other with the DD cord and fascia showing a similar relationship, the nodules cluster within its own space and can be separated from the rest using PC1 *vs.* PC2. B PCA score plot of PC2 *vs.* PC3 from patient DD2. The fat and SON displays greater separation compare to PC1 *vs.* PC2. The overlap between cord and fascia is still observed.

passage number. This substantiates the clear separation between the cultures of different DD phenotypes (DD-nodule, DD-cord and internal-control transverse palmar fascia) in the early passages (1 and 2), which evaporated as the passage number increased further.

Principal component analysis of footprint spectra (Experiment 1.5). No meaningful trends or separation could be observed using PCA on footprint data.

Metabolic profiling of cell-cultures from DD and normal palmar fascia

Multivariate Analysis on GC-MS data acquired from three patients (Experiment 2). Having established that early passages have a more reproducible phenotype the next stage was to assess sample (cord, nodule, control) variability. GC-TOF-MS analysis was performed on 27 samples, constructed from 3 different tissue phenotypes (nodule (n = 3), cord (n = 3) and control (transverse palmar fascia (n = 3)) cultured in triplicates. A total of 129 metabolite features were identified, of which 69 are known and reported in supplementary Table S3. The unknown metabolites are excluded. To identify the source of greatest variation within the combined and individual groups of data for all samples MVA was employed. The initial stage of the data analysis strategy was similar in strategy to FT-IR analysis: unsupervised exploratory data analysis using PCA was employed to discover any natural groups within the data and any outliers before pre-processing.

Fig. 7 shows the scores plot following ANOVA-PCA on DD nodules, DD cords and DD-control transverse palmar fascia. Separation between nodules & control appeared in PC1 while separation between cords and control appeared in PC2. The loadings were plotted and the variables with data points further away from the origin (zero) that had more significant contributions toward the separation, were examined. This procedure was used to compare relative metabolite concentrations in samples and to see whether peaks (metabolite IDs) from loading plots corresponded with those from univariate analysis. The significant results from the MVA and Wilcoxon rank test[35] are described in Table 1 in terms of the significantly dysregulated metabolites and the super families of pathways to which they contribute. For nodule samples compared with normal fascia, 11 metabolites were identified as significantly different. Of these, 4 unique and identified metabolites showed a decrease in DD nodules; cysteine, phenylalanine, leucine and aspartic acid. A further 7 metabolite features (peaks) changed significantly but remained unidentified. For DD cords compared with normal fascia, 9 metabolite peaks were identified as significantly different. Of these 1 unique and identified metabolites showed an increase in DD cords (glycerol-3-phosphate) while 2 metabolites (leucine and pantothenic acid) showed a decrease in DD cords. Some of these trends are shown in the box-whisker plots in Fig. 8A–F and their original

Table 1 Metabolite peaks observed to be statistically different when comparing the intracellular extracts detected for Dupuytren's disease (DD) and control. ↑ denotes an increase in metabolite concentration and ↓ denotes a decrease in metabolite concentration. MET ID number represents an unknown detected metabolite. Bold font and underline indicates significant metabolites. N = DD nodules, C = DD cord and F = internal controls; transverse palmar fascia

No	HMDB accession ID	Metabolite name	N *vs.* F *p*-value	C *vs.* F	C *vs.* N	Pathway
1	HMDB00687	L-Leucine	**0.031** ↓	**0.008** ↓	0.489	Amino acid metabolism
2		Sugar	0.625	**0.016** ↓	0.463	Carbohydrate metabolism
3	HMDB00126	Glycerol 3-Phosphate	0.25	**0.016** ↑	**0.039** ↑	Carbohydrate Metabolism
4	HMDB00210	Pantothenic acid	0.688	**0.055** ↓	0.489	Mechanism of cofactors and vitamins
5	HMDB00158	L-Tyrosine	0.313	0.148	**0.035** ↑	Amino acid metabolism
6		Sugar	**0.031** ↓	0.25	0.241	Carbohydrate metabolism
7	HMDB00159	L-Phenylalanine	**0.031** ↓	0.383	**0.03** ↑	Amino acid metabolism
8	HMDB00182	L-Lysine	0.219	0.383	**0.035** ↑	Amino acid metabolism
9	HMDB00574	L-Cysteine	**0.031** ↓	0.461	0.252	Amino acid metabolism
10	HMDB00182	L-Lysine	0.688	0.547	**0.035**	Amino acid metabolism
11	HMDB00191	L-Aspartic acid	**0.031** ↓	0.844	0.064	Amino acid metabolism
12	HMDB00574	L-Cysteine	**0.031** ↓	0.945	0.064	Amino acid metabolism
13		Sugar	0.063	0.645	**0.004** ↓	Carbohydrate metabolism
14	HMDB00182	L-Lysine	0.063	1	**0.002** ↑	Amino acid metabolism
15	Unknown ID	MET 100	**0.031**	0.461	0.127	—
16	Unknown ID	MET 29	**0.031**	0.313	0.762	—
17	Unknown ID	MET 58	**0.031**	0.25	0.389	—
18	Unknown ID	MET 21	**0.031**	0.461	0.715	—
19	Unknown ID	MET 77	0.063	**0.016**	0.147	—
20	Unknown ID	MET 96	**0.031**	0.688	0.389	—
21	Unknown ID	MET 126	0.625	0.375	**0.019**	—
22	Unknown ID	MET 102	0.25	0.375	**0.004**	—
23	Unknown ID	MET 121	0.063	**0.031**	1	—
24	Unknown ID	MET 61	0.563	**0.055**	0.978	—
25	Unknown ID	MET 60	0.313	**0.039**	0.421	—
26	Unknown ID	MET 17	0.156	0.383	**0.026**	—
27	Unknown ID	MET 74	**0.031**	0.945	0.761	—
28	Unknown ID	MET 128	0.563	**0.023**	0.073	—
29	Unknown ID	MET 94	1	0.125	**0.001**	—
30	Unknown ID	MET 16	0.156	0.844	**0.008**	—
31	Unknown ID	MET 7	1	1	**0.022**	—

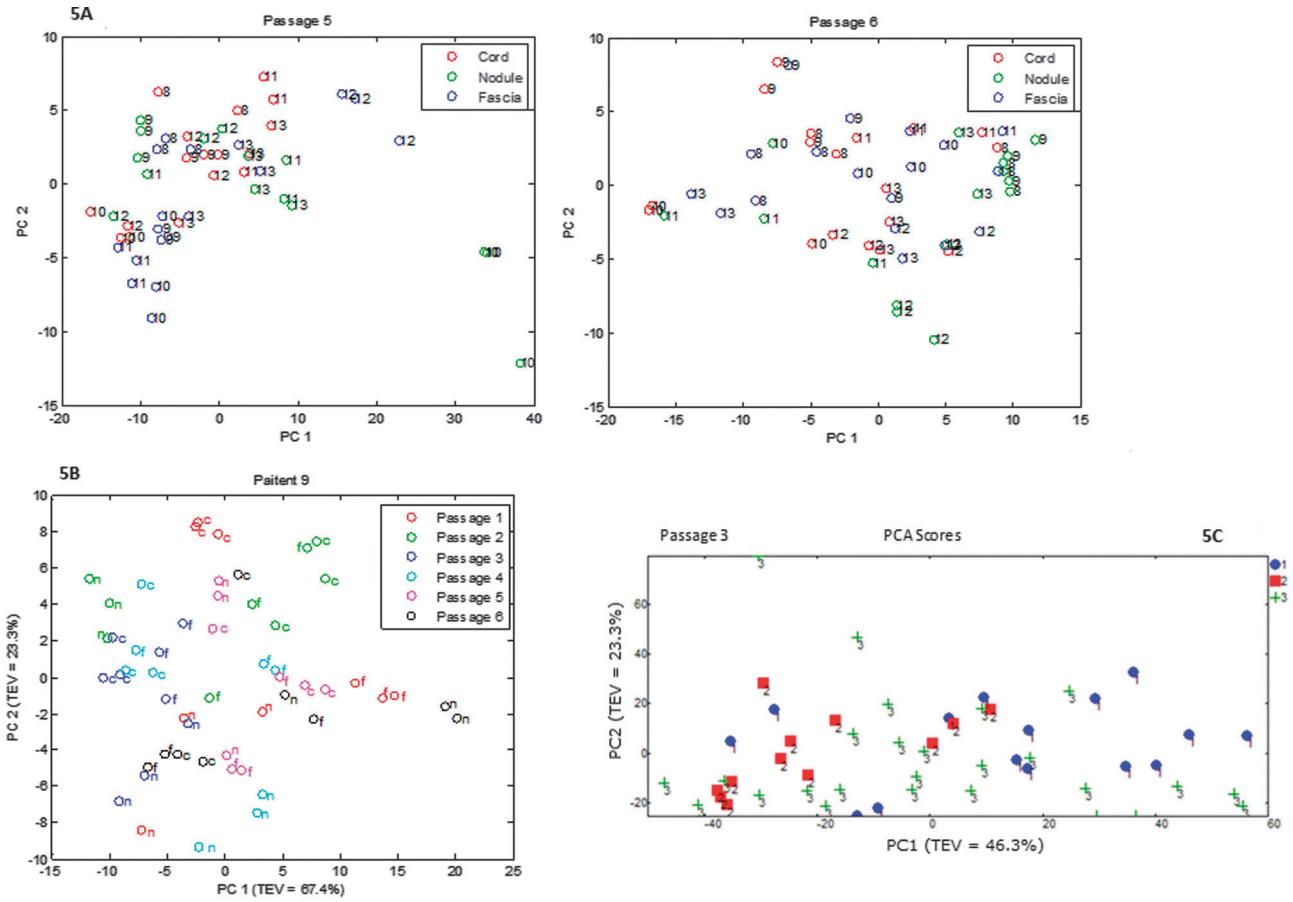

Fig. 5 A Projection of the Fourier transform infrared (FT-IR) spectra of Dupuytren's disease (DD) fibroblasts derived from the nodule, cord and fascia (control) with respect to passage number 5 and 6 onto the plane defined by PC1 and PC2. Patients are labeled by their numbers *e.g.* for each patients where the three cell types were analysed (*e.g.* DD nodule, cord and fascia), there are three different coloured circles. Red circle = cords, green circle = nodules, blue circle = fascia (internal control). B Projection of the FT-IR spectra of DD fibroblasts derived from the nodule, cord and fascia (control) with respect to their passage number onto the plane defined by PC1 and PC2 from a single patient. Samples from variable passage numbers are coloured. DD nodule, cord and fascia are labeled in lower case letters n, c, f respectively. C Projection of the FT-IR spectra of DD fibroblasts derived from the nodule, cord and fascia (internal control) with respect to passage number 3 onto the plane defined by t[1] (PC1) and t[2] (PC2). The samples are labeled by numbers and shapes; 1 blue circle = transverse palmar fascia, 2 red square = nodule, 3 green cross = cord.

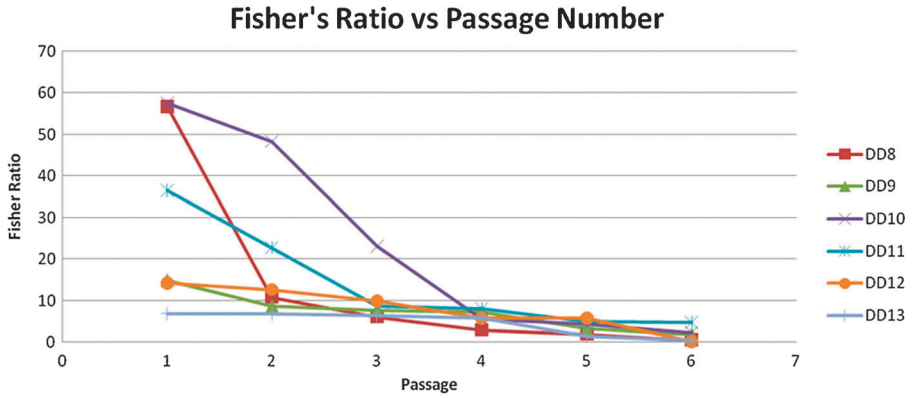

Fig. 6 Fisher ratio plot *vs.* passage number.

mass spectrometry spectra are provided in supplementary Fig. S5. A further 6 unidentified metabolite features were also significant.

MVA was also used to compare the DD cords against the DD nodules. From the Wilcoxon rank test,[35] 4 unique and identified metabolites showed a significant increase in cords compared to nodules; glycerol-3-phosphate, tyrosine, phenylalanine and lysine. A sugar was markedly down regulated in nodules. In addition 6 unidentified metabolites were also shown to have significant differences.

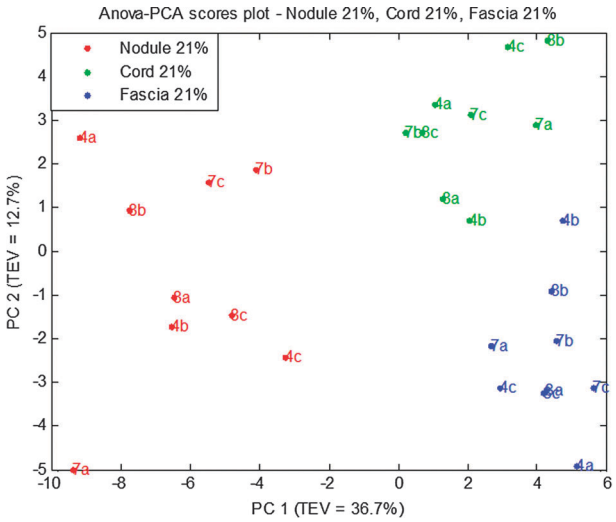

Fig. 7 ANOVA-PCA scores plot based on autoscaled gas chromatography-mass spectrometry (GC-MS) data showing the relationship between ANOVA-PCA scores plot based on autoscaled GC-MS data showing the relationship between nodules, cords and control. The numbers represent the patients. Each coloured dot/point represents a sample type. Number refers to patients in cohort C (14, 15 & 16) labeled 4, 7 and 8 respectively). Letters refer to replicates.

Pathway analysis. Analysis of all known metabolites in the dataset using MetPA[36] and KEGG database[37] revealed that members of the amino acid metabolism pathways were significantly overrepresented in the list of metabolites that changed in DD. Table 1 illustrates the directional change, with red arrows (up-regulated) and green arrows (downregulated). The results indicate that dysregulation in intermediates involved in carbohydrate and amino acid metabolism may attribute to DD formation. The dysregulation seems to correspond to rerouting rather than a general repression.

Discussion

A metabolomics phenotype for DD

The main aim of this study was to examine whether fibroblast cultures derived from different DD primary tissues from one hand and from two controls one from the same hand and the other from a healthy hand (patient unaffected with DD) exhibit differences at the metabolic level. The study showed unequivocally that they do. FT-IR spectroscopy not only discriminated DD phenotypes from the two controls, but also between the two fibrotic elements *i.e.* nodules from cords. Despite high biodiversity in patients, this trend could be observed in most patients when examined alone, if not altogether in the PCA scores plots when combined. Employing GC-MS this study identified metabolites that were significantly dysregulated in diseased tissue compared with the control fascia. The levels of four amino acids were significantly lower in DD nodules, *i.e.*, leucine, phenylalanine, cysteine, and aspartic acid. Leucine and an unidentified sugar molecule were also significantly down-regulated in DD cords as in DD nodules. Glycerol-3-phosphate (up-regulated) and pantothenic acid (down-regulated; the vitamin component of coenzyme A) were found to be up and down, respectively. They are involved in carbohydrate, energy and lipid metabolism. These findings can now support a more targeted profiling approach using LC-MS/MS. Two areas of metabolism are clearly highlighted for further systems-biology investigation. These include (a) amino acid (leucine, cysteine) metabolism and (b) carbohydrate/energy metabolism (CoA); moreover, the implications for fatty acid metabolism (CoA) should also be investigated. Variations in the intracellular metabolome reflect variations in biochemical reactions in cells and may aid hypothesis generation regarding the metabolic activities within DD cells as compared to control cells. From these variations, the pathways affected might then be inferred and disease pathogenesis or consequences better understood. In addition these findings for the metabolic level of cell function should be correlated with findings for other experimental findings notably the transcriptomic and the genomic data.

DD corresponds to permanent, progressive and irreversible contracture of tissue in the palm and digits. This could be due directly to an altered metabolism activating contractile elements, or to an altered anatomy of the tissue in terms of contractile proteins. Our observation of a metabolic phenotype is not yet sufficient evidence of the former possibility. Firstly, it is not clear how the metabolites identified would affect contractility directly. Secondly, the DD metabolic phenotype in cell culture was sufficiently long-lived to have to be co-determined by more stable phenotypes (the metabolite mentioned have turnover times below 1 min). Proteomic studies, in conjunction with metabolomics and systems biology, should be a most valuable next step.

The cell cultures are forgetful of the DD phenotype

Before interpreting any data related to physiological and pathological transitions in DD, it is important to determine "What is normal?" This also pertains to systematic variations in activity that accompany changes in cells and tissues as we move them from their natural environment to the context that enables us to analyse them: in this case "normal" is not a single phenotype, but a succession of adapting states. Although recent research in DD progression has taken a step closer to gaining insight into biologically related gene sets from the transverse palmar fascia,[10,12,38–40] none to date have explored the possibly dominant effects of sub-cultivation on any unique molecular or biochemical signature of DD. Few studies even state the passage number of the cells analysed.

Normal cells undergo only a finite number of divisions (a process known as replicative senescence), whereas tumour cells are able to proliferate indefinitely.[41,42] Passaging cell cultures requires the division of the cells, an activity the differentiated cells in and around DD tissue would not be much engaged in. Although different numbers of passages have been reported in some studies investigating Dupuytren disease, any change in the proliferative potential of the fibroblasts has not been assessed. DD fibroblasts may possess a higher potential for matrix and collagen production required to persist through passages than control fascia cells do because only the DD nodules and DD cords result from an uncontrolled proliferative state. Differences in collagen, fibronectin proteins, matrix expression proteins and even proteoglycans could be

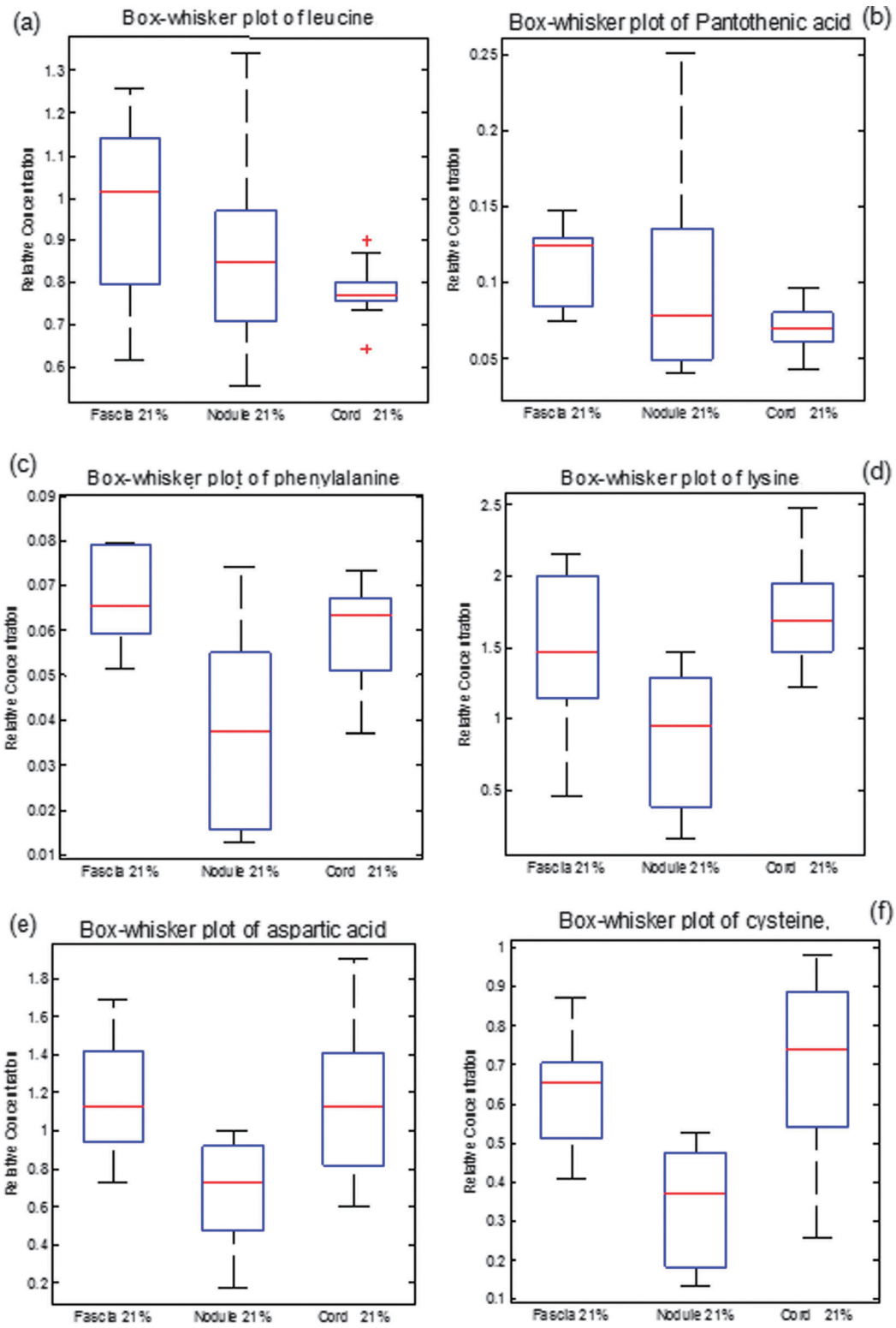

Fig. 8 Box-whisker plots demonstrating altered expression of metabolites in intracellular metabolomes of nodules, cords and Fascia cultured in 21% oxygen. (A) leucine, (B) pantothenic acid, (C) phenylalanine, (D) lysine, (E) aspartic acid, (F) cysteine. The red line in the box represents the median change in the peak value; the lower and upper boundaries of the box represent the 25th and 75th percentiles, respectively; the lower and upper whiskers represent the 5th and 95th percentiles, respectively; and crosses represent the outliers.**$p \leq 0.05$ Wilcoxon sign rank test.

affected by cell passage because at earlier passages all normal cells would mostly be proliferating while at later passages they would tend to senesce.[43]

Because of these and other considerations we hypothesized that when one brings DD and control cells in culture, they will change phenotype with the number of times they are passaged

into a new cell culture dish. We also expected that the difference between the phenotype of the DD cells and the control cells would fade away. Although the phenotype might reside mostly in the protein expression pattern, we expected that the metabolomics phenotype would reflect this. The results of this study support both hypotheses: The cell culture monolayer environment alters the functional characteristics of the cells with cell-passage number and the difference between DD and normal disappeared after 3 passages. It is unclear whether this is due to adaptation of the cells to their new, same environment or to selection for a subpopulation of cells which were already more prone to survive the *in vitro* conditions. 3D cultures may improve over cell monolayer culturing and should be considered in future studies.

Because DD comes with increased proliferative potential, one might conjecture that the disappearance of the difference between the metabolic phenotypes of DD and normal cells with passaging, was due to the normal phenotype moving to the DD phenotype and the latter remaining more or less constant.

The methodology

Within the biosciences, the applications of FT-IR have been numerous and diverse.[17,44,45] FT-IR spectroscopy can detect and identify endogenous (metabolic fingerprint) and secreted (metabolic footprint) metabolites.[17] Also in the present study FT-IR spectroscopy proved a powerful tool for the rapid screening and discrimination of fingerprints, between the anatomical cell types within individual DD patients. However, PCA of FT-IR footprint spectra did not yield any statistically meaningful results. This may be because the culture medium in which secreted metabolites were collected was nutrient rich and contained undefined reagents including FBS.

Much of the success of FT-IR spectroscopy derives from precise data analysis. Due to uneven sample size (and cellular content), projection analyses methods were used and PCA and PC-DFA indicated good clustering and separation between different sample types (from disease and healthy and also between samples from different individuals). Supervised analysis methodologies such as DFA and ANOVA-PCA were employed in addition to unsupervised methods (PCA) to make inferences from the mass spectral data.

A standard operation procedure in the search for DD mechanism?

We examined how DD and healthy cells changed following excision and cultivation over a period of time. The results from the metabolic fingerprinting based on PCA of DD nodule, cord and fascia imply that an early passage number cell (0–3) would faithfully represent the test subject as its phenotype would be closest to the *in vivo* state. We also identified principal components both in FT-IR and in GC-MS methodologies, as well as a number specific compounds as potential markers of the DD phenotype. Metabolic fingerprinting has the major advantages of speed, sensitivity, the ability to analyse many hundreds of samples simultaneously and the potential to generate a rich and highly informative data set. In addition FT-IR demonstrated that internal fascia can serve as an appropriate control and can be distinguished from diseased fibroblasts using our chemometric techniques. The use of internal transverse palmar fascia as the control will contribute to homogeneity and consistency in future studies.

Taken together this may begin to define a standard operating procedure for future studies after the mechanisms behind DD: DD cells put into cell culture, harvested after a first passage (in order to increase cell number) and checked for DD signature in terms of our FT-IR or GC-MS procedures. Transverse palmar fascia should be taken as the control.

Conclusions

Systematic hypothesis-driven studies have been performed to investigate the DD state dynamics *in vitro* with respect to passage numbers (*i.e.* over a time period). A number of metabolites dysregulated in DD have been identified. These metabolites are involved in amino acid metabolism, carbohydrate metabolism and include a vitamin required for the synthesis of an important coenzyme. Only cell cultures of early (0–3) passage numbers provide metabolic fingerprints representative of DD. This study has helped towards the elucidation of how genetic and environmental changes together disturb tissue function in DD and the results may encourage further research in optimally defined conditions with an emphasis on reproducibility and experimental robustness. Metabolites identified from these early studies may serve as the basis of larger, prospective, externally validated studies in clinical cohorts for their future use as potential diagnostic biomarkers.

List of abbreviations

ANOVA-PCA	Analysis of variance – principal component analysis
CTD	Carpal ligamentous fascia from carpel tunnel decompression
DD	Dupuytren's disease
DF1	Discriminant function 1
DF2	Discriminant function 2
DFA	Discriminant function analysis
DMEM	Dulbecco's Modified Eagle's Medium
DMSO	Dimethyl sulfoxide
DPBS	Dulbecco's phosphate buffered saline
EMSC	Extended multiplicative scatter correction
FBS	Fetal bovine serum
FDR	False discovery rate
FT-IR	Fourier transform-infrared
GC-MS	Gas chromatography-mass spectrometry
KEGG	Kyoto Encyclopedia of Genes and Genomes
MetPA	Metabolomics pathway analysis
MVA	Multivariate analyses
NEAA	Non-essential amino acids
PC-DFA	Principal component – discriminant function analysis
PCA	Principal component analysis
SB	Systems biology
SON	Skin overlying the nodule

Competing interests

The authors declare that they have no conflicting interests.

Authors' contributions

All authors contributed to manuscript preparation. SR performed sample preparation for all studies, acquired FT-IR spectral data and multivariate data analysis. WBD acquired GC-MS data. YX helped with acquisition of FT-IR spectral data and contributed to data analysis. PJRD contributed to the analysis of data. HW contributed to the design of the study and had a major involvement in the interpretation of the data in terms of the systems biology of DD, and in much of the formulation of the manuscript. AB conceived the study and obtained ethical approval for consenting and collection of tissue samples, contributed to the experimental design and the funding of the project, identified appropriate patients and tissue phenotypes. RG contributed to the design of the study and his major involvement is in the data generation of FT-IR and GC-MS as well as the multivariate data analyses.

Acknowledgements

We thank Ms Dulharie Wijeratne for her assistance in generating the histological samples and Prof. Anthony Freemont for kindly confirming the DD phenotype in histological samples. We thank Prof. Magnus Rattray for advisory aspects in this multidisciplinary project. SR thanks BBSRC and EPSRC for her studentship *via* the Doctoral Training Centre ISBML, at the Manchester Centre for Integrative Systems Biology, UK. This work was also supported by the EPSRC/BBSRC Manchester Centre for Integrative Systems Biology grant and the DTC grants to HW, as well as by other grants supporting his work from the EPSRC (EP/D508053/1), BBSRC (BB/C008219/1, BB/5302251, BB/I004688/1, BB/I017186/1), NWO (various), and EU-FP7 (EC-MOAN [NO043235], BioSim, NucSys [MRTN-CT-019496]). AB is in receipt of a personal award from the NIHR (UK).

References

1 S. Rehman, R. Goodacre, P. J. Day, A. Bayat and H. Westerhoff, Dupuytren's: a systems biology disease, *Arthritis Res. Ther.*, 2011, **13**, 238.
2 G. M. Rayan, Dupuytren Disease: Anatomy, Pathology, Presentation, and Treatment, *J. Bone Jt. Surg., Br. Vol.*, 2007, **89**, 189–198.
3 B. Shih and A. Bayat, Scientific understanding and clinical management of Dupuytren disease, *Nat. Rev. Rheumatol.*, 2010, **6**, 715–726.
4 I. T. H. Au-Yong, C. J. Wildin, J. J. Dias and R. E. Page, A review of common practice in Dupuytren surgery, *Techniques in Hand and Upper Extremity Surgery*, 2005, **9**, 178–187.
5 J. J. Tomasek, M. B. Vaughan and C. J. Haaksma, Cellular structure and biology of Dupuytren's disease, *Hand. Clin.*, 1999, **15**, 21–34.
6 G. Gabbiani and G. Majno, Dupuytren's contracture: fibroblast contraction? An ultrastructural study, *Am. J. Pathol.*, 1972, **66**, 131–146.
7 N. A. Hill and L. C. Hurst, Dupuytren's contracture, *Hand. Clin.*, 1989, **5**, 349–357.
8 R. Orsi, F. Zorzi, G. Brunelli and G. Sacchi, Histological and ultrastructural observations in 15 cases of Dupuytren's disease, *Osservazioni istologiche ed ultrastrutturali in 15 casi di morbo di Dupuytren*, 1983, **75**, 829–836.
9 R. H. Gelberman, D. Amiel, R. M. Rudolph and R. M. Vance, Dupuytren's contracture. An electronic microscopic, biochemical, and clinical correlative study, *Journal of Bone and Joint Surgery - Series A*, 1980, **62**, 425–432.
10 D. Ulrich, F. Ulrich, A. Piatkowski and N. Pallua, Expression of matrix metalloproteinases and their inhibitors in cords and nodules of patients with Dupuytren's disease, *Arch. Orthop. Trauma Surg.*, 2009, **129**, 1453–1459.
11 L. Satish, W. A. Laframboise and D. B. O'Gorman, Identification of differentially expressed genes in fibroblasts derived from patients with Dupuytren's Contracture, *BMC Med. Genomics*, 2008, **1**, 1–10.
12 S. Rehman, F. Salway, J. K. Stanley, W. E. R. Ollier, P. Day and A. Bayat, Molecular Phenotypic Descriptors of Dupuytren's Disease Defined Using Informatics Analysis of the Transcriptome, *J. Hand Surg.*, 2008, **33**, 359–372.
13 H. Westerhoff, A. Kolodkin, R. Conradie, S. Wilkinson, F. Bruggeman, K. Krab, J. van Schuppen, H. Hardin, B. Bakker and M. Moné, *et al.* Systems biology towards life *in silico*: mathematics of the control of living cells, *J. Math. Biol.*, 2009, **58**, 7–34.
14 P. Daran-Lapujade, S. Rossell, W. M. van Gulik, M. A. H. Luttik, M. J. L. de Groot, M. Slijper, A. J. R. Heck, J.-M. Daran, J. H. de Winde and H. V. Westerhoff, *et al.* The fluxes through glycolytic enzymes in Saccharomyces cerevisiae are predominantly regulated at posttranscriptional levels, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 15753–15758.
15 R. Goodacre, S. Vaidyanathan, W. B. Dunn, G. G. Harrigan and D. B. Kell, Metabolomics by numbers: acquiring and understanding global metabolite data, *Trends Biotechnol.*, 2004, **22**, 245–252.
16 W. B. Dunn, D. I. Broadhurst, H. J. Atherton, R. Goodacre and J. L. Griffin, Systems level studies of mammalian metabolomes: the roles of mass spectrometry and nuclear magnetic resonance spectroscopy, *Chem. Soc. Rev.*, 2011, **40**, 387–426.
17 D. I. Ellis and R. Goodacre, Metabolic fingerprinting in disease diagnosis: biomedical applications of infrared and Raman spectroscopy, *Analyst*, 2006, **131**, 875–885.
18 D. I. Broadhurst and D. B. Kell, Statistical strategies for avoiding false discoveries in metabolomics and related experiments, *Metabolomics*, 2006, **2**, 171–196.
19 W. B. Dunn, N. J. C. Bailey and H. E. Johnson, Measuring the metabolome: current analytical technologies, *Analyst*, 2005, **130**, 606–625.
20 G. D. Lewis, A. Asnani and R. E. Gerszten, *Application of Metabolomics to Cardiovascular Biomarker and Pathway Discovery*, 2008, vol. 52, pp. 117–123.
21 L. M. Raamsdonk, B. Teusink, D. Broadhurst, N. Zhang, A. Hayes, M. C. Walsh, J. A. Berden, K. M. Brindle, D. B. Kell and J. J. Rowland, *et al.* A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations, *Nat. Biotechnol.*, 2001, **19**, 45–50.
22 J. R. Bain, Sr, B. R. Wenner, O. Ilkayeva, D. M. Muoio and C. B. Newgard, Metabolomics Applied to Diabetes Research: Moving From Information to Knowledge, *Diabetes*, 2009, **58**, 2429–2443.
23 H. L. T. Wu, C. Ma, R. Xue, C. Deng, H. Zeng and X. Shen, GC/MS-based metabolomic approach to validate the role of urinary sarcosine and target biomarkers for human prostate cancer by microwave-assisted derivatization, *Anal. Bioanal. Chem.*, 2011, **401**, 635–646.
24 W. B. Dunn, R. Goodacre, L. Neyses and M. M, Integration of metabolomics in heart disease and diabetes research: current achievements and future outlook, *Bioanalysis*, 2011, **3**, 2205–2222.
25 F. Tom, *Extended multiplicative scatter correction*, 2005, vol. 16, pp. 3–5.
26 R. M. Jarvis, D. Broadhurst, H. Johnson, N. M. O'Boyle and R. Goodacre, *PYCHEM: a multivariate analysis package for python*, 2006, vol. 22, pp. 2565–2566.
27 S. Wold, K. Esbensen and P. Geladi, Principal component analysis, *Chemom. Intell. Lab. Syst.*, 1987, **2**, 37–52.
28 H. Wold, in *Multivariate Analysis*, Academic Press, 1966.
29 B. F. Manly, *Multivariate statistical methods: a primer*, J Chapman & Hall, London, United Kingdom 1994.
30 P. Begley, S. Francis-McIntyre, W. B. Dunn, D. I. Broadhurst, A. Halsall, A. Tseng, J. Knowles, R. Goodacre and D. B. Kell, Development and Performance of a Gas Chromatography Time-of-Flight Mass Spectrometry Analysis for Large-Scale Nontargeted Metabolomic Studies of Human Serum, *Anal. Chem.*, 2009, **81**, 7038–7046.

31 W. B. Dunn, D. Broadhurst, P. Begley, E. Zelena, S. Francis-McIntyre, N. Anderson, M. Brown, J. D. Knowles, A. Halsall and J. N. Haselden, *et al*. Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry, *Nat. Protoc.*, 2011, **6**, 1060–1083.

32 L. W. Sumner, A. Amberg, D. Barrett, M. H. Beale, R. Beger, C. A. Daykin, T. W. M. Fan, O. Fiehn, R. Goodacre and G. JL, *et al*. Proposed minimum reporting standards for chemical analysis, *Metabolomics*, 2007, **3**, 211–221.

33 M. Brown, W. B. Dunn, P. Dobson, Y. Patel, C. L. Winder, S. Francis-McIntyre, P. Begley, K. Carroll, D. Broadhurst and A. Tseng, *et al*. Mass spectrometry tools and metabolite-specific databases for molecular identification in metabolomics, *Analyst*, 2009, **134**, 1322–1332.

34 R. Goodacre, D. Broadhurst, A. Smilde, B. S. Kristal, J. D. Baker, R. Beger, C. Bessant, S. Connor, G. Capuani and A. Craig, *et al*. Proposed minimum reporting standards for data analysis in metabolomics, *Metabolomics*, 2007, **3**, 231–241.

35 F. Wilcoxon, Individual Comparisons by Ranking Methods, *Biometrics Bull.*, 1945, **1**, 80–83.

36 J. Xia and D. S. Wishart, MetPA: a web-based metabolomics tool for pathway analysis and visualization, *Bioinformatics*, 2010, **26**, 2342–2344.

37 M. Kanehisa, S. Goto, Y. Sato, M. Furumichi and M. Tanabe, KEGG for integration and interpretation of large-scale molecular data sets, *Nucleic Acids Res.*, 2012, **40**, D109–D115.

38 B. Shih, J. Brown, D. Armstrong, T. Lindau and A. Bayat, Differential Gene Expression Analysis of Subcutaneous Fat, Fascia, and Skin Overlying a Dupuytren's Disease Nodule in Comparison to Control Tissue, *HAND*, 2009, **4**, 294–301.

39 B. Shih, D. Wijeratne, D. J. Armstrong, T. Lindau, P. Day and A. Bayat, Identification of Biomarkers in Dupuytren's Disease by Comparative Analysis of Fibroblasts Versus Tissue Biopsies in Disease-Specific Phenotypes, *J. Hand Surg.*, 2009, **34**, 124–136.

40 H. Seyhan, J. Kopp, S. Schultze-Mosgau and R. E. Horch, Increased Metabolic Activity of Fibroblasts Derived from Cords Compared with Nodule Fibroblasts Sampling from Patients with Dupuytren's Contracture, *Plast. Reconstr. Surg.*, 2006.

41 C. Schöneich, Proteomics in gerontological research, *Exp. Gerontol.*, 2003, **38**, 473–481.

42 J. K. Leung and O. M. Pereira-Smith, *Identification of Genes Involved in Cell Senescence and Immortalization: Potential Implications for Tissue Ageing*, in *Ageing Vulnerability: Causes and Interventions*, ed. Gregory Bock and J. A. Goode, 2003, pp. 105–115.

43 S. C. R. Benvenuti, J. Bruce, M. D. Waterfield and P. S. Jat, Identification of novel candidates for replicative senescence by functional proteomics, *Oncogene*, 2002, **21**(28), 4403–4413.

44 D. Naumann, D. Helm and H. Labischinski, Microbiological characterizations by FT-IR spectroscopy, *Nature*, 1991, **351**, 81–82.

45 R. Goodacre, E. M. Timmins, R. Burton, N. Kaderbhai, A. M. Woodward, D. B. Kell and P. J. Rooney, Rapid identification of urinary tract infection bacteria using hyperspectral whole-organism fingerprinting and artificial neural networks, *Microbiology*, 1998, **144**, 1157–1170.