AUTOIMMUNITY

Synovial fibroblast gene expression is associated with sensory nerve growth and pain in rheumatoid arthritis

Zilong Bai[1], Nicholas Bartelo[1], Maryam Aslam[2], Elisabeth A. Murphy[2], Caryn R. Hale[2,3], Nathalie E. Blachere[2,4], Salina Parveen[2], Edoardo Spolaore[5], Edward DiCarlo[5], Ellen M. Gravallese[6], Melanie H. Smith[5], Accelerating Medicines Partnership RA/SLE Network†‡, Mayu O. Frank[2], Caroline S. Jiang[2], Haotan Zhang[1], Christina Pyrgaki[2], Myles J. Lewis[7], Shafaq Sikandar[7], Costantino Pitzalis[7,8], Joseph B. Lesnak[9], Khadijah Mazhar[9], Theodore J. Price[9], Anne-Marie Malfait[10], Rachel E. Miller[10], Fan Zhang[11], Susan Goodman[5], Robert B. Darnell[2,4], Fei Wang[1]*, Dana E. Orange[2,5]*

It has been presumed that rheumatoid arthritis (RA) joint pain is related to inflammation in the synovium; however, recent studies reveal that pain scores in patients do not correlate with synovial inflammation. We developed a machine-learning approach (graph-based gene expression module identification or GbGMI) to identify an 815-gene expression module associated with pain in synovial biopsy samples from patients with established RA who had limited synovial inflammation at arthroplasty. We then validated this finding in an independent cohort of synovial biopsy samples from patients who had early untreated RA with little inflammation. Single-cell RNA sequencing analyses indicated that most of these 815 genes were most robustly expressed by lining layer synovial fibroblasts. Receptor-ligand interaction analysis predicted cross-talk between human lining layer fibroblasts and human dorsal root ganglion neurons expressing calcitonin gene–related peptide (CGRP$^+$). Both RA synovial fibroblast culture supernatant and netrin-4, which is abundantly expressed by lining fibroblasts and was within the GbGMI-identified pain-associated gene module, increased the branching of pain-sensitive murine CGRP$^+$ dorsal root ganglion neurons in vitro. Imaging of solvent-cleared synovial tissue with little inflammation from humans with RA revealed CGRP$^+$ pain-sensing neurons encasing blood vessels growing into synovial hypertrophic papilla. Together, these findings support a model whereby synovial lining fibroblasts express genes associated with pain that enhance the growth of pain-sensing neurons into regions of synovial hypertrophy in RA.

INTRODUCTION

Inflammatory pain can be driven by cytokines, bradykinins, and prostanoids, which bind specific receptors on primary nociceptor neurons to cause heightened sensation of pain (*1*). However, pain is not always proportional to inflammation, and clinical scenarios in which pain is dissociated from inflammation are useful to study the noninflammatory drivers of pain.

Rheumatoid arthritis (RA) is a chronic disease characterized by inflammation in the synovium, the tissue that lines the joint cavity. Remarkable progress has been made in developing an array of conventional synthetic, targeted synthetic, and biologic disease-modifying antirheumatic drugs (csDMARDs, tsDMARDs, and bDMARDs, respectively), which target relevant immune mediators (*2*). However, up to 20% of patients with RA are "difficult to treat"; that is, they do not improve despite treatment with at least two bDMARDs or tsDMARDs, with different mechanisms of action, after failing a csDMARD (*3*, *4*). It has been assumed that synovial inflammation is the cause of RA joint pain. However, recent studies have revealed that pain can be dissociated from inflammation in RA (*3*–*8*). Patients with RA and limited synovial inflammation, also known as "fibroid," "low inflammatory," "pauci-immune," or "fibroblast cell type abundance phenotype" synovium, have as much pain as those with extreme inflammation (*3*, *4*, *9*–*13*). Patients with low synovial inflammation tend to receive less benefit from treatment with anti-inflammatory drugs such as tumor necrosis factor inhibitors and DMARDs (*12*, *14*).

Here, we hypothesized that a focused analysis of low inflammatory synovium might identify factors beyond inflammation that relate to joint pain. One challenge is that patient-reported outcome data, such as report of pain severity, are notoriously noisy, and tissue transcriptomic data are still relatively expensive, with only tens to hundreds of samples available for most studies. Given these limitations, existing analytic approaches are generally not sufficiently powered to identify gene modules that relate to patient reported outcomes. To address this, we developed a machine-learning approach, called graph-based gene expression module identification (GbGMI), to uncover the relationship between gene expression and pain. We then sought to determine which cells express pain-associated genes and how they might influence nociceptors in vitro and to validate these findings by visualizing neurons in low inflammatory RA synovium in vivo.

[1]Weill Cornell Medicine, New York, NY 10065, USA. [2]Rockefeller University, New York, NY 10065, USA. [3]Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA. [4]Howard Hughes Medical Institute, Rockefeller University, New York, NY 10065, USA. [5]Hospital for Special Surgery, New York, NY 10021, USA. [6]Brigham and Women's Hospital, Boston, MA 02115, USA. [7]Queen Mary University of London & NIHR BRC Barts Health NHS Trust, London E1 4NS, UK. [8]Department of Biomedical Sciences, Humanitas University & IRCC Humanitas Research Hospital, Milan 20072, Italy. [9]University of Texas at Dallas, Richardson, TX 75080, USA. [10]Rush University Medical Center, Chicago, IL 60612, USA. [11]University of Colorado School of Medicine, Aurora, CO 80045, USA.
*Corresponding author. Email: few2001@med.cornell.edu (F.W.); dorange@rockefeller.edu (D.E.O.)
†These authors contributed equally to this work.
‡Accelerating Medicines Partnership RA/SLE Network collaborators and affiliations appear at the end of this paper.

RESULTS

Pain is not related to inflammation in patients with RA with low inflammatory synovium

We categorized patients as high or low inflammatory using our previously reported histology scoring algorithm (*9*). RA pain scores were not different between patients with high and low inflammatory synovium (Fig. 1A). Pain scores were associated with the grade of synovial inflammation as measured by the density of cells per unit of tissue (in cells per square millimeter) in patients with high inflammatory synovium but not in patients with low inflammatory synovium (Fig. 1B).

We tested for genes that were highly associated with pain using the usual RNA sequencing (RNA-seq) analysis platform, limma (*15*). We failed to identify any individual genes that were correlated with pain, suggesting that the relationship of gene expression with pain could be multifactorial or nonlinear. We next hypothesized that there might be groups of genes whose expression varies in association with pain. We developed an iterative machine-learning GbGMI computational framework to uncover a group of genes whose expression is correlated with a given univariate clinical feature. GbGMI was given a multimodal input comprising a gene expression matrix M for m genes and n patients and an n-dimensional clinical feature vector $\boldsymbol{a}$. GbGMI first calculates the patient-to-patient similarity structure according to the given clinical feature and compares that to the gene expressions using the Laplacian score. GbGMI then determines the optimal number of genes that together associate with the clinical feature through statistical tests between the t-distributed stochastic neighborhood embedding (t-SNE)–based summary scores of the selected genes and this clinical feature (see Materials and Methods and fig. S1).

GbGMI correctly identifies genes associated with synovial inflammation

We sought to test the validity of the approach by testing whether it would correctly identify genes known to be associated with inflammation as measured by cell density, which is highly associated with many individual genes as measured by limma (*15*). GbGMI identified a module of 2713 genes whose gene expression summary score correlates with synovial tissue cell density (Fig. 2, A to D). The positive control for this analysis was principal component one (PC1) of bulk synovial RNA-seq gene expression data, which was previously shown to associate with the extent of synovial inflammation and highly correlate with synovial cell density (*16*). The negative control was a gene expression summary score for a group of the top 5000 most variably expressed genes. As expected, the gene expression summary score of the top 5000 most variably expressed genes was not significantly correlated with synovial histologic cell density ($P = 0.21$) (Fig. 2E), whereas PC1 scores of gene expression were significantly correlated with synovial histologic cell density (Spearman $\rho = 0.4$, $P = 0.01$) (Fig. 2F). The gene expression summary score of the GbGMI module of 2713 genes had a further improved correlation to synovial histologic cell density (Spearman $\rho = 0.59$, $P = 0.0001$) (Fig. 2G). This analysis indicated that GbGMI was a useful method that outperformed principal components analysis (PCA) in identifying a module of genes that associate with the grade of synovial inflammation measured by cell density.

GbGMI identifies pain-associated synovial gene expression in patients with established RA

We next applied GbGMI to define a module of genes associated with pain in patients with low inflammatory synovium. Most of the 6582 genes that distinguish high and low inflammatory synovium are increased in high inflammatory synovium and are enriched for pathways representing infiltrating immune cells. To uncover genes associated with pain but not inflammation, we focused our analysis on 2227 genes that exhibited increased expression in low inflammatory synovium relative to high inflammatory synovium (*9*) and on pain scores that document the extent of pain in the joint that was sampled [Hip Osteoarthritis Outcome Score/Knee Osteoarthritis Outcome Score (HOOS/KOOS)] (Fig. 3A). The patient-reported pain scores a were transformed into a matrix of pairwise similarity scores between patients S (Fig. 3B). We next calculated the Laplacian score (*17*) for each of the 2227 low inflammatory genes on the basis of its expression values (that is, a row vector in M) and S (Fig. 3C). We then tested which number of top-ranked genes collectively best correlated with pain among patients with RA with low synovial inflammation and identified an 815-gene module, which we refer to as the GbGMI-identified pain-associated genes (Fig. 3D). Although the summary score of all 2227 low inflammatory genes did not correlate with pain, summary scores of the GbGMI-identified pain-associated genes were significantly correlated with the patient-reported HOOS/KOOS pain in patients with

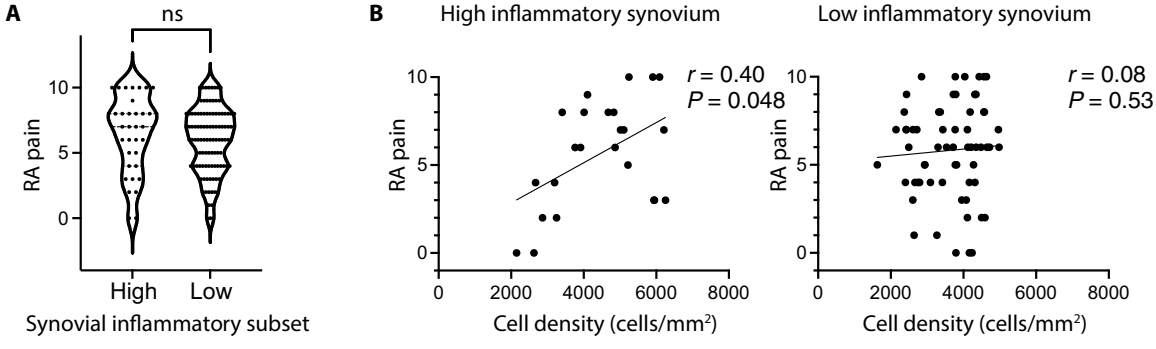

Fig. 1. Pain is related to synovial inflammation in patients with RA with high, but not low, synovial inflammation. (**A**) RA pain scores are shown compared to synovial tissue inflammatory classification in $n = 139$ patients. (**B**) RA pain scores are shown according to cell density (in cells per square millimeter) of H&E (hematoxylin and eosin)–stained synovial tissue, in samples classified as high ($n = 35$, $r = 0.40$, $P = 0.048$) or low inflammatory ($n = 104$, $r = 0.08$, $P = 0.53$). ns, not significant in Mann-Whitney test. r = Spearman's rank correlation coefficient. P = two-tailed P value.

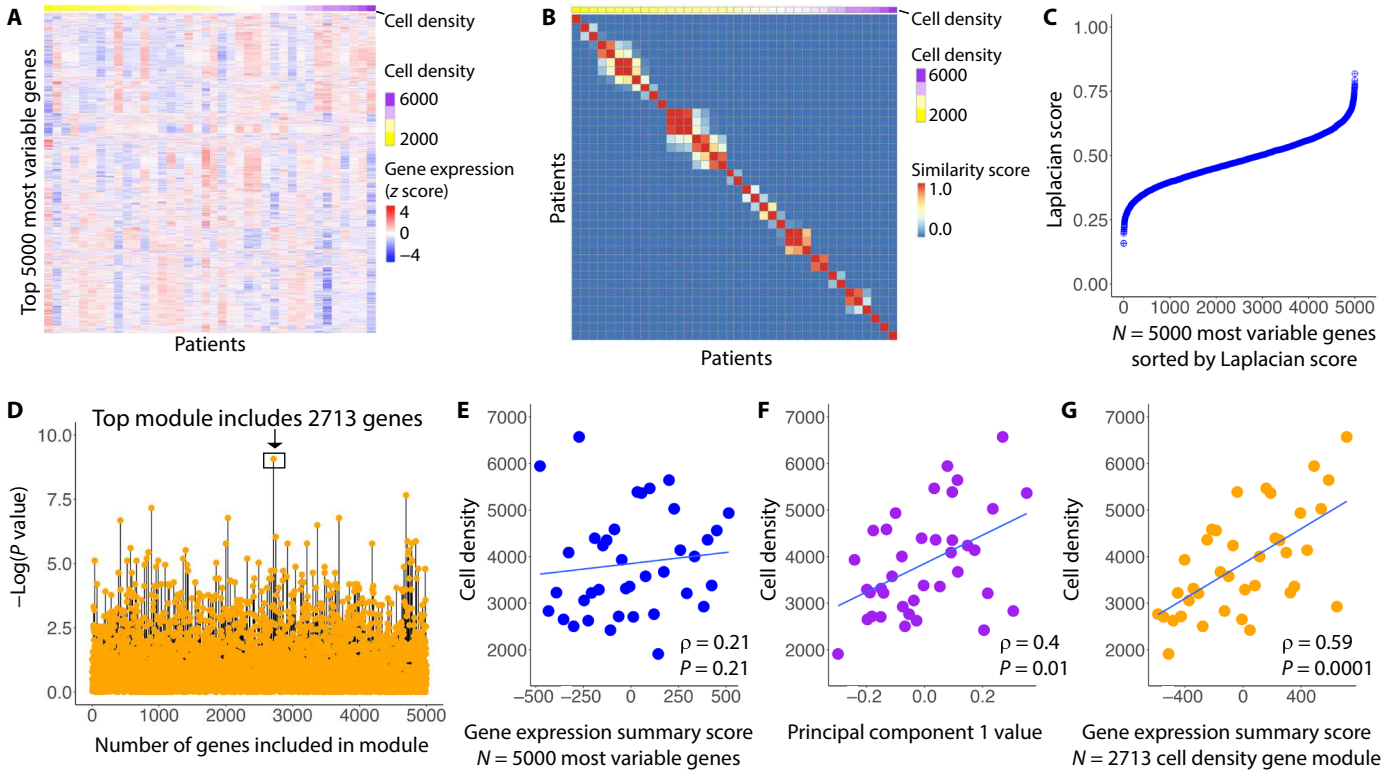

Fig. 2. A synovial gene expression signature that correlates with synovial histologic cell density was identified using GbGMI or PCA. (**A**) An expression heatmap is shown of the top 5000 most variably expressed genes in 38 patients. Gene expression amounts (rows) are represented as *z* scores for all patients. Patients (columns) are sorted by their mean nuclei densities. (**B**) Shown is a similarity matrix of synovial histologic cell densities. (**C**) Laplacian scores are shown for the top 5000 most variably expressed genes measuring how their expression varied compared with synovial histologic cell density similarity structure. Each dot represents a gene, sorted by Laplacian score in ascending order. (**D**) Shown is the −log(*P* value) of the correlation of the top *k*–ranked groups of genes with nucleus density similarity structure. Each dot represents a group of genes. (**E**) Synovial histologic cell density is shown according to PC1 score of the top 5000 most variably expressed genes in 38 patients (Spearman $\rho = 0.21$, $P = 0.21$). (**F**) Synovial histologic cell density is shown according to the summary score of the 5000 genes over the 38 patients (Spearman $\rho = 0.4$, $P = 0.01$). (**G**) Synovial histologic cell density is shown according to the summary score of the 2713 GbGMI-identified genes over the 38 patients (Spearman $\rho = 0.59$, $P = 0.0001$). Statistics presented in (E) to (G) indicate Spearman correlation coefficient and *P* value.

low inflammatory synovium ($P = 0.001$) (Fig. 3E). This correlation was not as pronounced when including all patients with RA irrespective of inflammatory subset (Fig. 3F). Similar correlations were identified when the GbGMI-identified pain-associated genes were compared to Visual Analog Score (VAS) report of pain (fig. S2). We conducted sensitivity analysis on GbGMI using 22 patient subsets subsampled with a leave-one-out strategy. Despite the relatively small sample size, GbGMI demonstrated robustness; on these subsampled data, GbGMI yielded pain-association gene rankings consistent with the original ranking based on the overall 22 patients with low inflammation (smallest $\rho = 0.75$, largest $P = 0$, Spearman's test; fig. S3A) and identified pain-associated gene modules significantly associated with the original 815-gene module (largest $P = 1.22 \times 10^{-7}$, Fisher's exact test; fig. S3B, last row and column).

GbGMI-identified pain-associated synovial genes are also associated with pain in patients with early RA

Overfitting is a concern in using a graph-based machine learning approach to identifying groups of genes that associate with pain. It is possible that the GbGMI-identified pain-associated genes correlated with pain in the dataset in which they were found but not in other external datasets. We sought to test whether the pain-associated gene module identified in patients with established disease was also associated with pain in a second, independent Pathobiology of Early Arthritis Cohort (PEAC) dataset (*11*) of synovial biopsy samples from patients with early (mean of 6 months of symptoms) untreated RA. A total of 2018 of the 2227 low inflammatory genes and 738 of the 815 pain-associated genes found in the established RA dataset were also detected in this dataset. The 738 pain-associated genes were also correlated with VAS pain in patients with early RA with low inflammatory (fibroid or undefined) synovium (Fig. 3G). However, the 2018 low inflammatory genes were not. In this early RA cohort, the 738 genes were also associated with pain when samples from all patients were included as well, irrespective of synovial inflammatory subset (fibroid, undefined, myeloid, and lymphoid), although the association was again not as robust as was seen in those with low inflammatory synovium (Fig. 3H). The range of GbGMI summary scores decreased when all samples were included (Fig. 3, G and H). The association of the GbGMI-identified genes with pain was robust in the low inflammatory samples but persisted even when all patients were included, suggesting that these genes may play a role, albeit less pronounced, in pain in high inflammatory synovium, where inflammatory mediators are likely to also contribute.

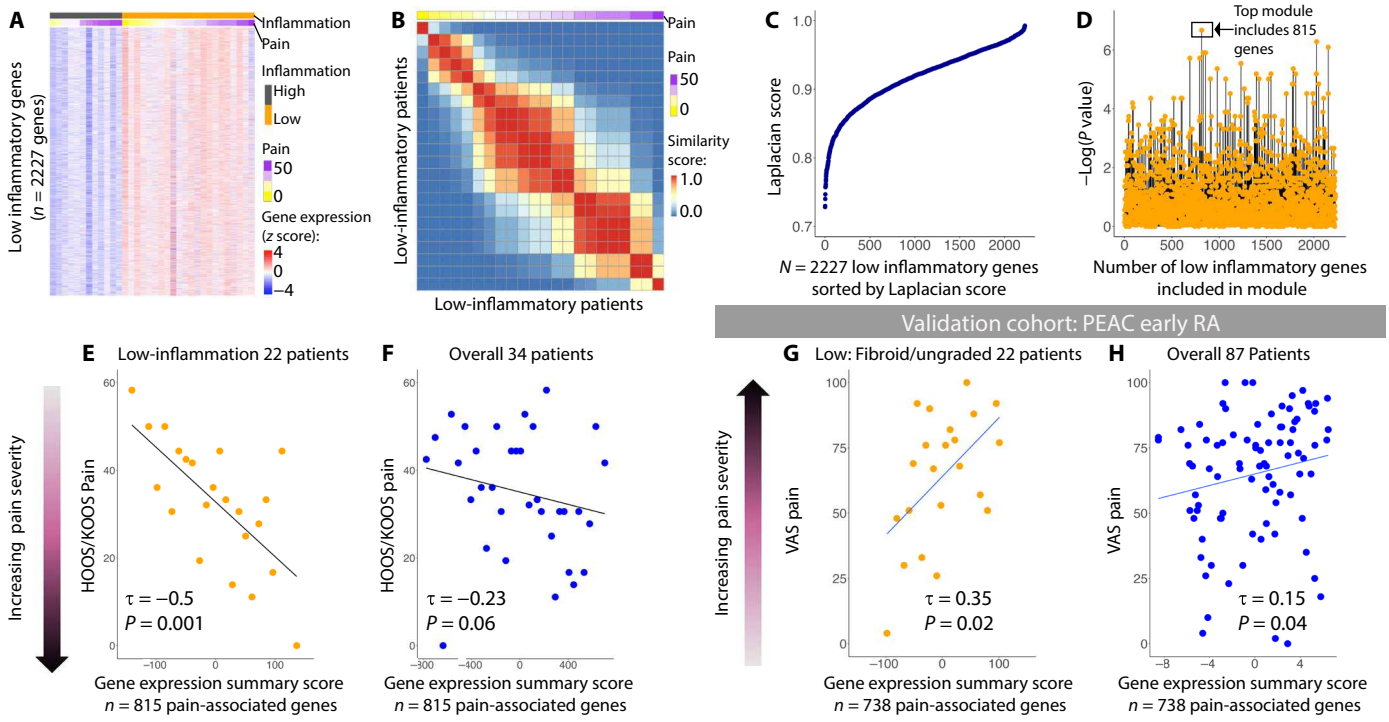

Fig. 3. GbGMI identified a pain-associated synovial gene expression in patients with established and early RA. (**A**) An expression heatmap is shown of 2227 genes with increased expression in low inflammatory synovium. Patients (columns) were grouped by inflammatory amounts: high ($n = 12$) versus low ($n = 22$, low and mixed inflammatory subtypes identified in (*9*). (**B**) Similarity structure is shown of HOOS/KOOS pain scores over patients. (**C**) Laplacian scores of the input 2227 genes measure how their expression amounts over patients relate to the pain score–based similarity structure. Each dot represents a gene sorted by Laplacian score in ascending order. (**D**) Significance of the correlation between the top *k*–ranked genes and HOOS/KOOS pain scores is shown. Each dot represents a group of genes. (**E**) Established RA HOOS/KOOS pain score according to the summary score of the 815 GbGMI-identified genes in low inflammatory samples (Kendall $\tau = -0.5$, $P = 0.001$). (**F**) Comparison between the same pair of scores as (E) but for all samples, irrespective of inflammatory status (Kendall $\tau = -0.23$, $P = 0.06$). (**G**) Early RA VAS pain scores are shown according to gene expression summary score of 738 GbGMI-identified pain-associated genes in patients with low inflammatory (fibroid or unassigned) synovial pathotype (Kendall $\tau = 0.35$, $P = 0.02$). (**H**) Comparison between the same pair of scores as (G) but for all patients, irrespective of synovial pathotype (Kendall $\tau = 0.15$, $P = 0.04$). *z* score was calculated from gene expression values over patients. Statistics presented in (E) to (H) indicate Kendall correlation coefficient and *P* value.

GbGMI-identified pain-associated genes are enriched with neurogenesis pathways and predominantly expressed by synovial fibroblasts

We next sought to understand the biological meaning and the direction of the association of the 815 GbGMI genes with pain in patients with RA with low synovial inflammation. Limma was performed to detect genes whose expression correlated with pain. Genes were ranked by limma according to this correlation. Although limma did not identify any significant [false discovery rate (FDR) < 0.05] individual genes correlated with pain (fig. S4A), as a group, expression of the 815 GbGMI-identified pain genes was significantly decreased as the HOOS/KOOS pain score increased (adjusted $P = 7.38 \times 10^{-12}$, Kolmogorov-Smirnov test) (fig. S4, B and C). This indicated a positive correlation with pain severity. The 815 pain-associated genes were enriched for pathways such as nervous system development, neurogenesis, and neuron differentiation (Fig. 4A) and included ephrin (*EPHA3*, *EPHA6*, and *EPHA7*) and semaphorin (*SEMA3B*, *SEMA3E*, *SEMA4C*, *SEMA5A*, and *SEMA6D*) family members. The 1412 nonpain-associated genes included *CD55*, *PRG4*, *CSPG4*, and *MERTK*, genes known to be involved in the normal functions of lining macrophages and fibroblasts (*18*). These genes were enriched in molecular function and ribosomal RNA (rRNA) processing but not neuron axonal growth pathways (Fig. 4A and data file S1). We next examined which cells expressed the GbGMI pain-associated genes. We compared expression in both bulk RNA-seq data from sorted cell types, which offers high-depth coverage of RNA but less cell type resolution, and single-cell RNA-seq (scRNA-seq) data, which offers higher cell type resolution to cell subtypes but less depth of coverage, from the Accelerating Medicines Partnership dataset (*19*). Comparison of the pain-associated genes across sorted bulk synovial B cells, fibroblasts, monocytes, and T cells indicated that the fibroblasts exhibited the highest expression of pain-associated genes (Fig. 4, B and C). We reasoned that pain-associated genes might be more robustly expressed in fibroblasts because of a relative enrichment in fibroblasts, compared with immune cells, in low inflammatory samples. However, when looking only at fibroblasts, the pain-associated genes were increased compared with the nonpain-associated genes (Fig. 4B). The differences in pathways enriched in pain-associated and nonpain-associated genes as well as the difference in relative expression within fibroblasts indicated that the GbGMI method did not select a random group of fibroblast genes. Further analysis of the scRNA-seq dataset also confirmed that the fibroblast subsets exhibited the highest expression of pain-associated genes (Fig. 4D). Gene expression analysis among fibroblast subsets indicated that, compared with the other fibroblast subsets, lining $CD55^+$ fibroblasts (SC-F4)

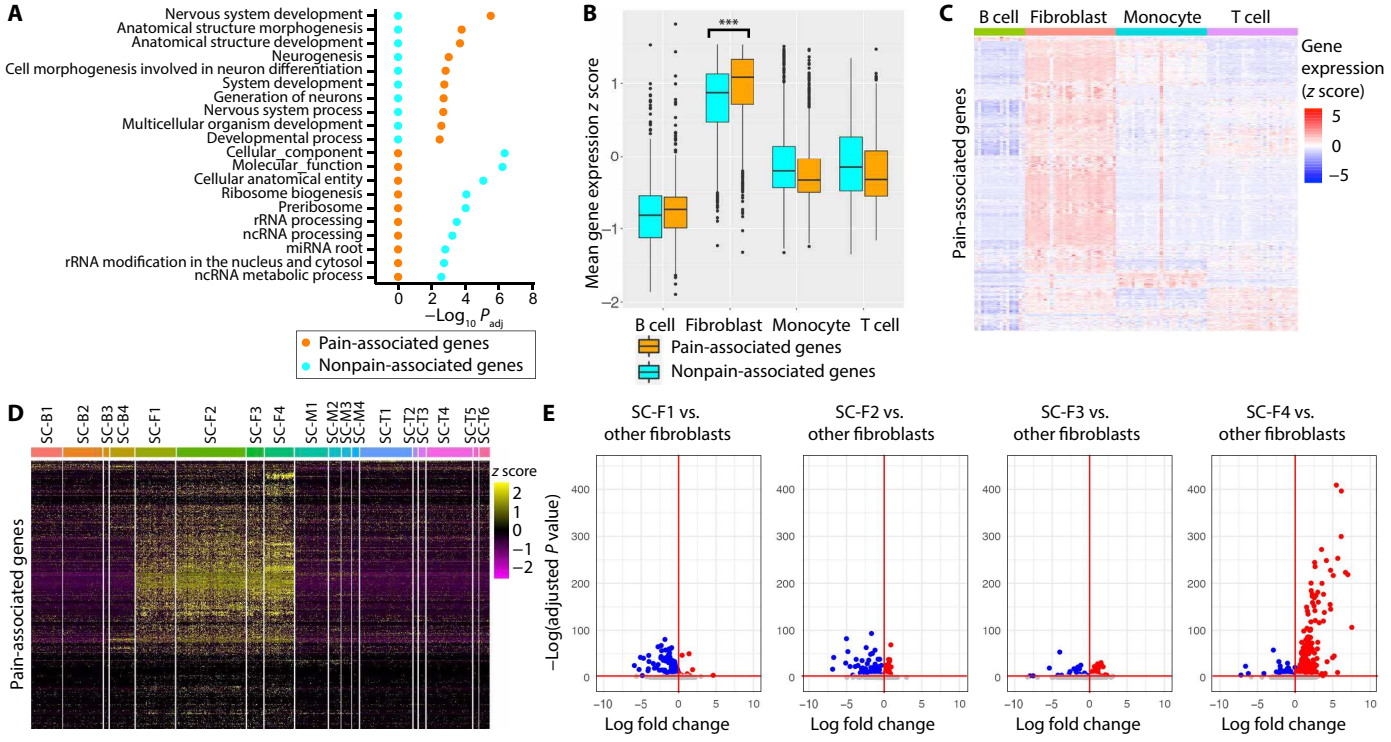

Fig. 4. The GbGMI-identified pain-associated gene signature is expressed by synovial lining layer fibroblasts. (**A**) g:Profiler pathway enrichment analysis is shown of 815 pain-associated genes and 1412 nonpain-associated low inflammatory genes. ncRNA, long noncoding RNA; miRNA, microRNA. (**B**) Mean gene expression *z* scores of pain-associated genes and nonpain-associated genes detected in sorted bulk B cells ($CD45^+$, $CD3^-$, and $CD19^+$), fibroblasts ($CD45^-$, $CD31^-$, and $PDPN^+$), monocytes ($CD45^+$ and $CD14^+$), and T cells ($CD45^+$ and $CD3^+$) are shown. *z* score is calculated on the basis of TPM normalized counts. ***$P < 0.001$. (**C**) Per-sample gene expression *z* scores of 769 pain-associated genes detected in sorted cell types from (B) are shown. (**D**) An expression heatmap is shown of 797 pain genes with nonzero variance in expression values across a subset ($n = 4354$) of RA synovial cells (SC) in 18 unique cell populations (of B cells, SC-B1 to SC-B4; fibroblasts, SC-F1 to SC-F4; monocytes, SC-M1 to SC-M4; and T cells, SC-T1 to SC-T6), which were identified from the 5265 scRNA-seq profiles by an integrated analysis based on canonical-correlation analysis (CCA) from the Accelerating Medicine Partnership (*19*). *z* score is calculated on the basis of $\log_2$(CPM + 1)–transformed UMI counts over the RA synovial cells. (**E**) Volcano plots of 794 pain genes in scRNA-seq profiles (ImmPort accession #SDY998) (*19*) with nonzero variance in expression values across the subset ($n = 1532$) of RA synovial fibroblasts in three sublining subsets, $CD34^+$ (SC-F1), HLA-DRA^{hi} (SC-F2), and $DKK3^+$ (SC-F3), and one lining subset (SC-F4) are shown. Each volcano plot shows the differential expression analysis (using Seurat function FindMarkers) of the genes in each RA synovial fibroblast subtype compared with the other three, where *x* axis shows $\log_2$(fold change) and *y*-axis −log(adjusted *P* value). The significantly increased genes are red, significantly decreased genes are blue, and nonsignificantly differentially expressed genes are gray. The horizontal and vertical red lines respectively indicate the threshold of significance [−log(adjusted $P = 0.05$)] and the separation threshold between increased and decreased gene expression $\log_2$(fold change) = 0.

exhibited the highest expression of GbGMI-identified pain-associated genes (Fig. 4E).

Ligand-receptor analysis predicts interactions between lining fibroblasts and human dorsal root ganglion nociceptors

Given that the pain-associated genes were enriched in neuron axonal growth pathways, we next explored predicted interactions of pain-associated synovial fibroblast genes with dorsal root ganglion (DRG) neurons likely to innervate the joint. We performed receptor-ligand interaction analysis to identify predicted receptor-ligand pairs using the pain-associated genes expressed by four synovial fibroblast subtypes in human RA synovial tissue and genes expressed in a human DRG (hDRG) single-nucleus RNA-seq (snRNA-seq) dataset (Fig. 5A) (*20*). Lining fibroblasts (SC-F4) were predicted to have the highest number of ligand-receptor interactions (39 SC-F4 ligands to hDRG receptors) (fig. S5A). Lining fibroblasts (SC-F4) were predicted to interact with several $CGRP^+$ peptidergic and non-peptidergic neuron subtypes (fig. S5, B and C, and data file S2). Comparison of the expression of 14 ligand or receptor encoding pain-associated genes of SC-F4 revealed a gradient of pain-associated genes that were relatively lowly expressed in SC-F1 cells and most highly expressed in SC-F4 cells, with *HBEGF*, *CTGF*, and *NTN4* among the most robustly expressed (Fig. 5B).

We repeated this receptor-ligand interaction analysis using a bulk hDRG spatial RNA-seq dataset with neuronal expression confirmed in independently generated human snRNA-seq dataset, offering increased depth of coverage and a larger number of human samples from organ donors of both sexes (*21*–*23*) and a mouse DRG scRNA-seq dataset (*24*). From the 815 GbGMI fibroblast pain-associated genes, 158 unique ligand-receptor interactions were identified with receptors present on hDRG neurons (data file S3). The top 25 unique ligand-receptor interactions from synovial fibroblasts to hDRGs were ranked on the basis of the gene's Laplacian score (Fig. 5C). This demonstrated possible activation of several receptors on hDRGs involved in axon growth, including integrins (*ITGA5*, *ITGAM*, *ITGB2*, and *ITGB5*) and neogenin (*NEO1*) that is targeted by netrins. Biological process and molecular function Gene

Ontology (GO) terms were identified for the hDRG receptors in the interactome using Enrichr (*25*, *26*). This demonstrated activation of hDRG receptors responsible for axon guidance, nervous system development, and dendrite function (data file S4). Pathway analysis (iPathwayGuide software; https://advaitabio.com/ipathwayguide) on hDRG receptor genes was used to determine biological pathways that could be activated in hDRG neurons by fibroblast pain-associated genes. This again revealed activation of several biological pathways within hDRGs, with axon guidance being the most enriched driven heavily by ephrin A and B signaling (Fig. 5D). This analysis also revealed enrichment of extracellular matrix receptor interactions, regulation of actin cytoskeleton, focal adhesion, and Ras-associated protein 1 (RAP1) signaling, all of which point to axon growth. To further explore possible cross-talk between DRGs and synovial

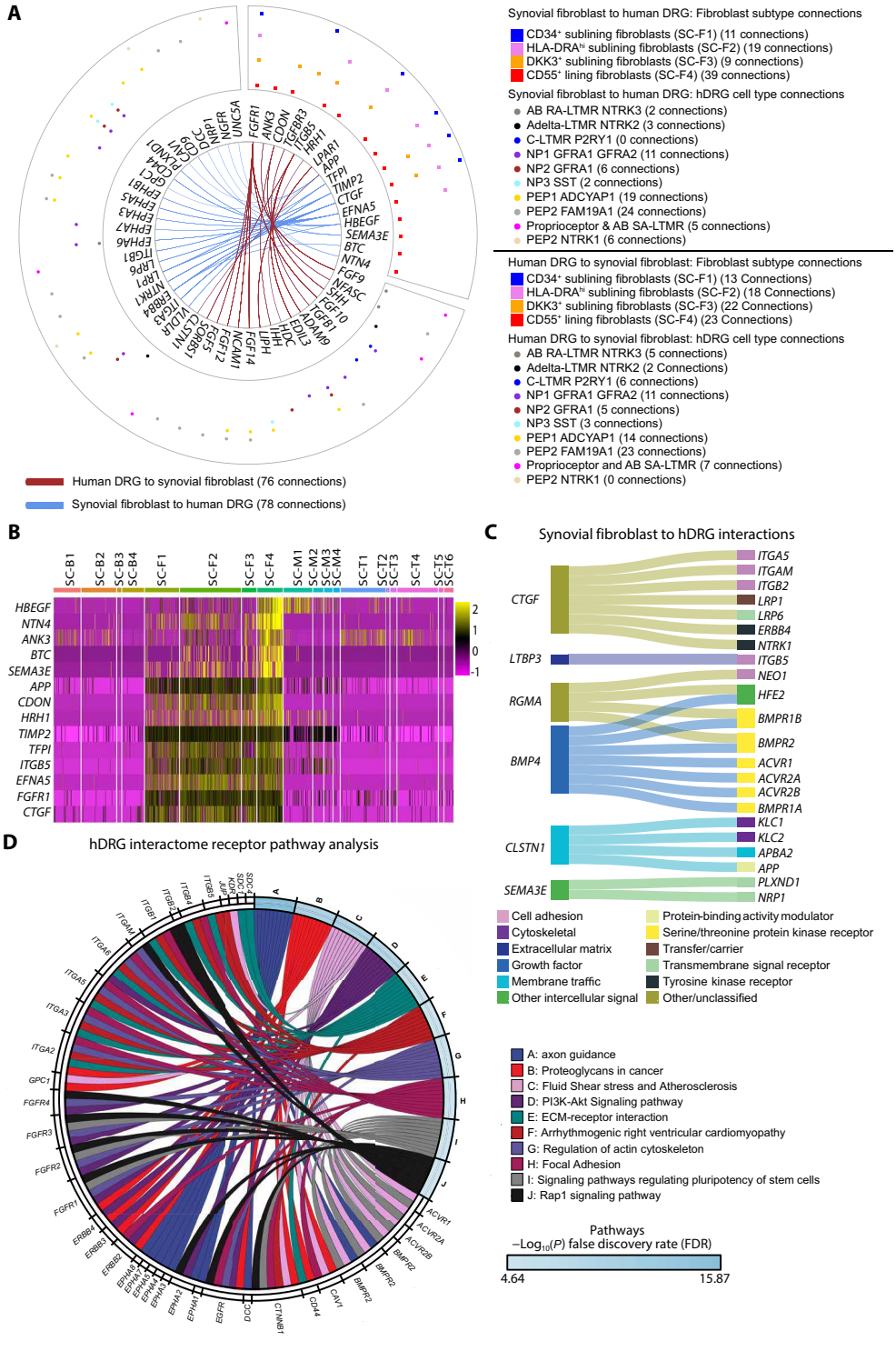

Fig. 5. Filtering on synovial fibroblast genes predicted to influence DRG sensory neurons. (**A**) Shown are predicted ligand-receptor interactions between synovial fibroblasts and neurons in the human DRG (hDRG). The circos plot shows the two unidirectional interactomes between the four synovial fibroblast subtypes and 10 hDRG neuron subtypes. The outermost layer indicates the RA synovial fibroblast subtypes of the cells (in colored squares) or hDRG neuron subtypes (in colored round dots) expressing corresponding ligand or receptor genes. The middle layer shows whether a gene is ligand coding or receptor coding in its associated interactions. The inner layer contains gene names. The two tissue-wise directions are distinguished by the colors of connections between gene names. The number of connections associated with the ligand/receptor genes in each fibroblast subtype or neuron subtype and those in each unidirectional tissue-wise relation are summarized in the corresponding legends. (**B**) An expression heatmap is shown of 14 pain-associated ligand/receptor encoding marker genes of synovial lining fibroblast (SC-F4) cells with nonzero variance in expression values across a subset (n = 4354) of RA synovial cells in 18 unique cell populations (of B cells, SC-B1 to SC-B4; fibroblasts, SC-F1 to SC-F4; monocytes, SC-M1 to SC-M4; and T cells, SC-T1 to SC-T6), which were identified from the 5265 scRNA-seq profiles by an integrated analysis based on CCA from the Accelerating Medicine Partnership (*19*). z score was calculated using $\log_2$(CPM + 1) transformed UMI counts over the RA synovial cells. The genes are ranked top-down by their log fold change in differential expression analysis (DEA) of lining fibroblast (SC-F4) versus other fibroblasts. (**C**) A Sankey plot is shown of top 25 unique fibroblast to hDRG ligand-receptor interactions from the 815 GbGMI pain-associated genes ranked by Laplacian score value. (**D**) A chord plot is shown depicting pathway analysis of hDRG receptors identified by the ligand receptor interactome by GbGMI pain-associated genes. The top enriched pathways suggest promotion of axon growth, including axon guidance, extracellular matrix (ECM)–receptor interaction, regulation of actin cytoskeleton, and Rap1 signaling.

fibroblasts, we also performed a ligand-receptor interactome between ligands from hDRG neurons and GbGMI synovial fibroblast pain-associated genes that encode for a receptor. This revealed 132 unique interactions between ligands from hDRGs and receptors from synovial fibroblasts (data file S5). Plotting the top 25 interactions ranked by Laplacian score (fig. S5D) revealed activation of protein tyrosine phosphatase receptors (*PTPRS*) and transforming growth factor–β (*TGFBR1*), both of which play roles in mediating cell growth and differentiation. GO term analysis of the fibroblast receptors revealed activation chemokine/cytokine signaling, immune cell migration, and fibroblast growth factor binding (data file S4).

Products of synovial fibroblasts influence adult DRG sprouting and branching in response to injury

We next sought to test whether any of the pain-associated synovial fibroblast genes found in this analysis might directly influence the growth of nociceptors in the synovium. $CGRP^+$ nerve fibers have been previously identified in synovium (*27*). *NTN4* was of interest because it was within the GbGMI-identified pain-associated gene module in this dataset, was highly expressed by synovial fibroblasts, and was identified in the hDRG interactome (data file S2). Although netrin-4 (NTN4) has only 30% sequence homology to NTN1, which plays a role in axon guidance during embryogenesis (*28*), NTN4 has been shown to augment embryonic olfactory bulb sprouting and thalamocortical branching (*29*–*31*). We cultured adult mouse–dissociated DRG neurons with no supplements (medium alone); human nerve growth factor (huNGF), as a positive control, because it is known to produce robust effects on axon sprouting (*32*, *33*); or human NTN4 (huNTN4) and measured survival, sprouting, and branching of $CGRP^+$ DRG neurons. CGRP status of the DRG neurons was assigned using immunofluorescent stains (fig. S6). The percentage of neurons with sprouting and the numbers of branches were quantified using Sholl analysis (Fig. 6A). The medium control in this experiment was Neurobasal Plus medium containing B27, which contained the protein transferrin. RA synovial fibroblasts express transferrin, but it was not included in the GbGMI-identified pain gene list, and its expression does not correlate with patient report of pain (fig. S7A) and therefore served as an internal control. Whereas there was no effect of either huNGF or huNTN4 on $CGRP^+$ neuron survival compared to $CGRP^+$ neurons cultured in medium control, huNTN4 augmented sprouting and branching in vitro (Fig. 6B). This result was reproduced when mouse neurons were cultured with mouse Ntn4 (fig. S7B), which has over 90% homology to huNTN4. However, synovial fibroblasts secrete an array of many different factors predicted to interact with nociceptors, and, considering the complexity of neuron axon guidance and the multitude of factors known to attract or repel neuron growth, we reasoned that it would be useful to determine whether the sum of all factors secreted by synovial fibroblasts resulted in nociceptive neuron attraction, repulsion, or neither. We cultured $CGRP^+$ neurons in RA synovial fibroblast–conditioned medium. Compared with medium alone, RA synovial fibroblast–conditioned medium augmented $CGRP^+$ neuron sprouting and branching (Fig. 6C).

$CGRP^+$ axons accompany blood vessels in RA synovial papillary hypertrophic processes

We next sought to delineate whether the structure of nociceptive neurons in the synovium of humans with RA are altered in vivo. We performed immunolabeling-enabled three-dimensional imaging of solvent-cleared organs (iDISCO) (*34*). Synovial papillary hypertrophy is a common feature of RA synovium, and blood vessels can be seen growing into these processes by gross inspection of unstained tissue (Fig. 7A) and hematoxylin and eosin (H&E)–stained synovial sections (Fig. 7B). SC-F4 fibroblasts, the cell type with the most robust expression of genes associated with patient report of pain end enriched for neural guidance pathway genes, line these papillary processes (*35*). Anti-CD31 antibody staining of cleared RA synovium revealed a profuse tangle of blood vessels growing into each papillary outgrowth (Fig. 7C and movies S1 and S2). High-magnification three-dimensional views inside the papillary processes revealed that $CGRP^+$ axons accompanied many of the blood vessels that had grown into the papillary processes (Fig. 7D). CGRP staining was just outside the CD31 stain, consistent with a meshwork of $CGRP^+$ axons wrapping around blood vessels (Fig. 7E), a neurovascular pattern previously observed with scanning electron microscopy (*36*). The presence of $CGRP^+$ axons in papillary processes of a synovial pannus indicated abnormal axonal sprouting in vivo because papillary processes represent abnormal synovial growth. We conclude that neoangiogenesis into abnormal papillary processes toward lining fibroblasts in the low inflammatory RA synovium was accompanied by neoneurogenesis of $CGRP^+$ nociceptive axons.

DISCUSSION

Recent studies have shown that RA synovial tissue spans a spectrum of synovial pathotypes, from highly inflamed synovium to fibroblast-enriched synovium (*13*). Here, we show that patients with RA can have relatively limited synovial inflammation but highly abnormal synovium with extensive papillary hypertrophy, angiogenesis, and high severity of pain. We demonstrated that synovial lining fibroblast products can augment the growth of $CGRP^+$ nociceptive axons/neurites and $CGRP^+$ axons extend into synovial papillary processes, placing nociceptors spatially adjacent to lining fibroblasts. These observations support a model whereby altered synovial lining fibroblasts secrete factors, including NTN4, that facilitate growth of $CGRP^+$ axons/neurites (fig. S8). These data are consistent with recent studies in osteoarthritis (OA), which identified aberrant $CGRP^+$ axonal sprouting into normally aneural cartilage (*37*–*40*).

Whereas RA tends to affect all three knee compartments similarly, OA tends to affect the medial side more severely. Compared with synovial fibroblasts from the nonpainful side, synovial fibroblast–conditioned medium from the painful side of knee OA increased neuron survival and longest branch length in vitro (*41*). Our studies also led to the discovery that NTN4 augments branching of injured pain-sensitive $CGRP^+$ nerves in vitro. NTN4 binds the extracellular matrix molecule, laminin, specifically laminin γ1, and, in doing so, markedly weakens matrix stiffness (*30*). Laminins are large glycoproteins that are abundant in synovium, and it is possible that a compliant extracellular matrix facilitates $CGRP^+$ axon sprouting or branching. NTN4 also binds cell surface receptors, such as neogenin (*42*), which are expressed by human nociceptors (*21*, *22*). Avenues for future research include in vivo animal studies to test the function of NTN4, as well as other pain-associated genes, and exploration of their potential as therapeutic targets.

The GbGMI approach is a combination of feature selection and data dimensionality reduction that enabled us to uncover a group of genes that together change their expression in association with

patient report of pain. Reducing the dimensionality of the data allowed us to detect associations with variable clinical data that were likely underpowered when all genes were included in the analysis as individual tests. Other data dimensionality reduction approaches, such as clustering and PCA, have been used in a similar way to detect clinical associations that would not be measurable in a gene-by-gene analysis (*9*). One key difference between the GbGMI approach and hierarchical clustering or PCA is that the latter groups

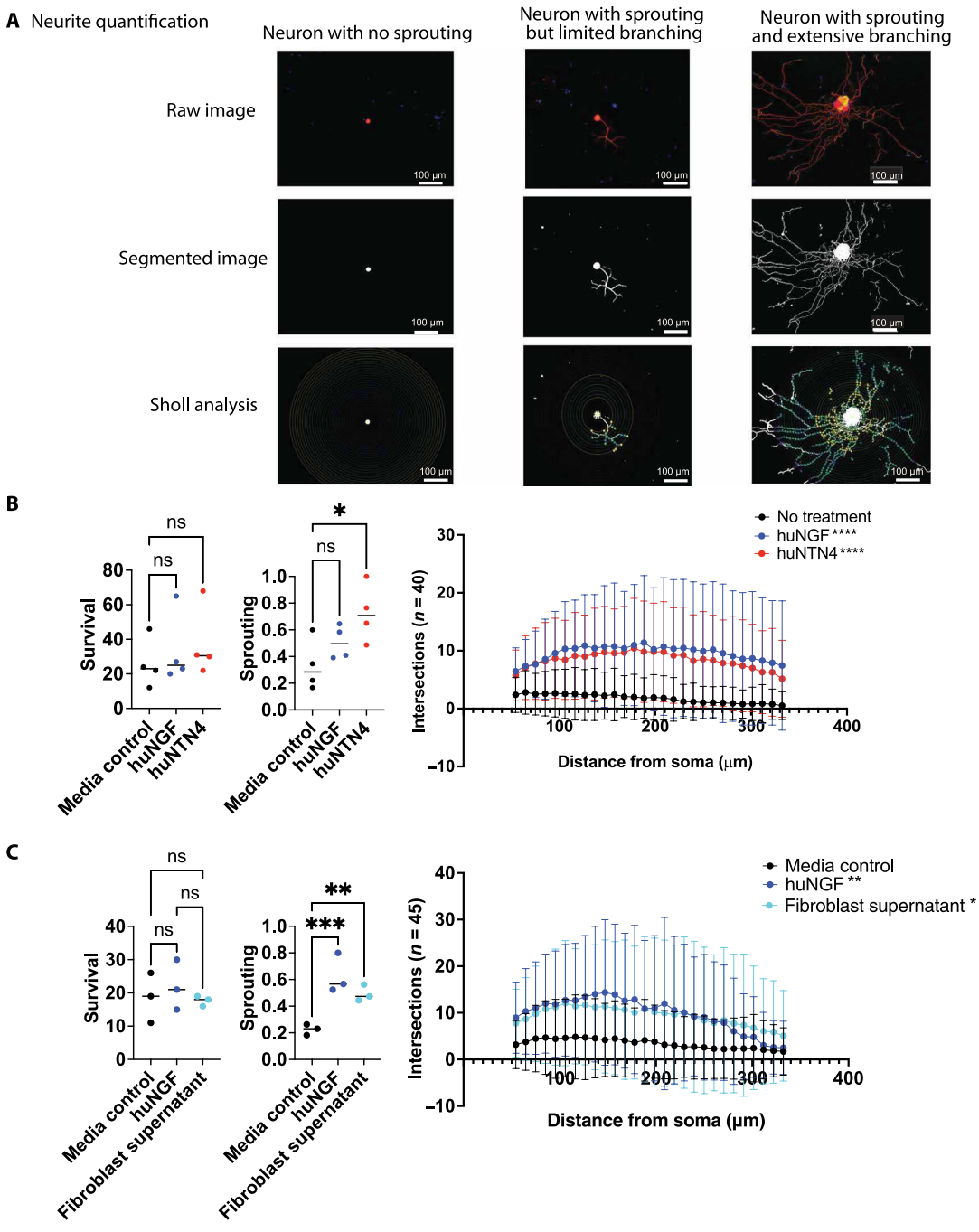

Fig. 6. Synovial fibroblast products affect DRG neurons in vitro. (**A**) Shown are representative images of the method used for neurite quantification of sprouting and Sholl analysis of branching of dissociated DRG neurons. (**B**) Quantification of CGRP$^+$ DRG neurons cultured with medium alone, huNGF, or huNTN4 is shown. Left: Survival was measured by the number of Map2b$^+$B3tub$^+$ cells of >10 µm. Each dot represents the sum of five ×10 magnification views from one experiment. Data from four experiments are presented. ns indicates not significant in Kruskal-Wallis test. Middle: The sum of sprouting neurons divided by the total number of neurons cultured with medium alone, huNGF, or huNTN4 is shown. Neurons with at least three axon branches greater than two times the size of the soma were classified as sprouting. Each dot represents the sum of five ×10 magnification views from one experiment. Data from four experiments are presented. *$P < 0.05$, ns indicates not significant in Kruskal-Wallis test. Right: Branching is shown as measured by the number of shell intersections of neurites, in DRG neurons cultured with medium alone (no treatment), huNGF, or huNTN4. Each dot represents the median with confidence interval of 40 neurons imaged from four experiments (10 neurons per experiment). ****$P < 0.0001$ in two-way ANOVA group*radius interaction with post hoc Dunnett's multiple comparisons of each treatment group with the no treatment group. (**C**) Quantification of CGRP$^+$ DRG neurons cultured with medium alone, medium supplemented with huNGF, or RA synovial fibroblast–conditioned medium is shown. Left: Survival was measured by the number of Map2b$^+$B3tub$^+$ CGRP$^+$ cells of >10 µm. Each dot represents the sum of five ×10 magnification views from one experiment. Data from three experiments are presented. ns indicates not significant in Kruskal-Wallis test. Middle: The sum of sprouting neurons divided by the total number of neurons cultured with medium alone, medium supplemented with huNGF, or RA synovial fibroblast–conditioned medium is shown. Neurons with at least three axon branches greater than two times the size of the soma were classified as sprouting. Each dot represents the sum of five ×10 magnification views from one experiment. Data from three experiments are presented. **$P < 0.01$ and ***$P < 0.001$ in Kruskal-Wallis tests. Right: Branching is shown as measured by the number of shell intersections of neurites in DRG neurons cultured with medium alone, medium supplemented with huNGF, or RA synovial fibroblast–conditioned medium. Each dot represents the median with confidence interval of 45 neurons imaged from three experiments (15 neurons per experiment). *$P < 0.05$ and **$P < 0.01$ in two-way ANOVA group*radius interaction with post hoc Dunnett's multiple comparisons of each treatment group with the no treatment group.

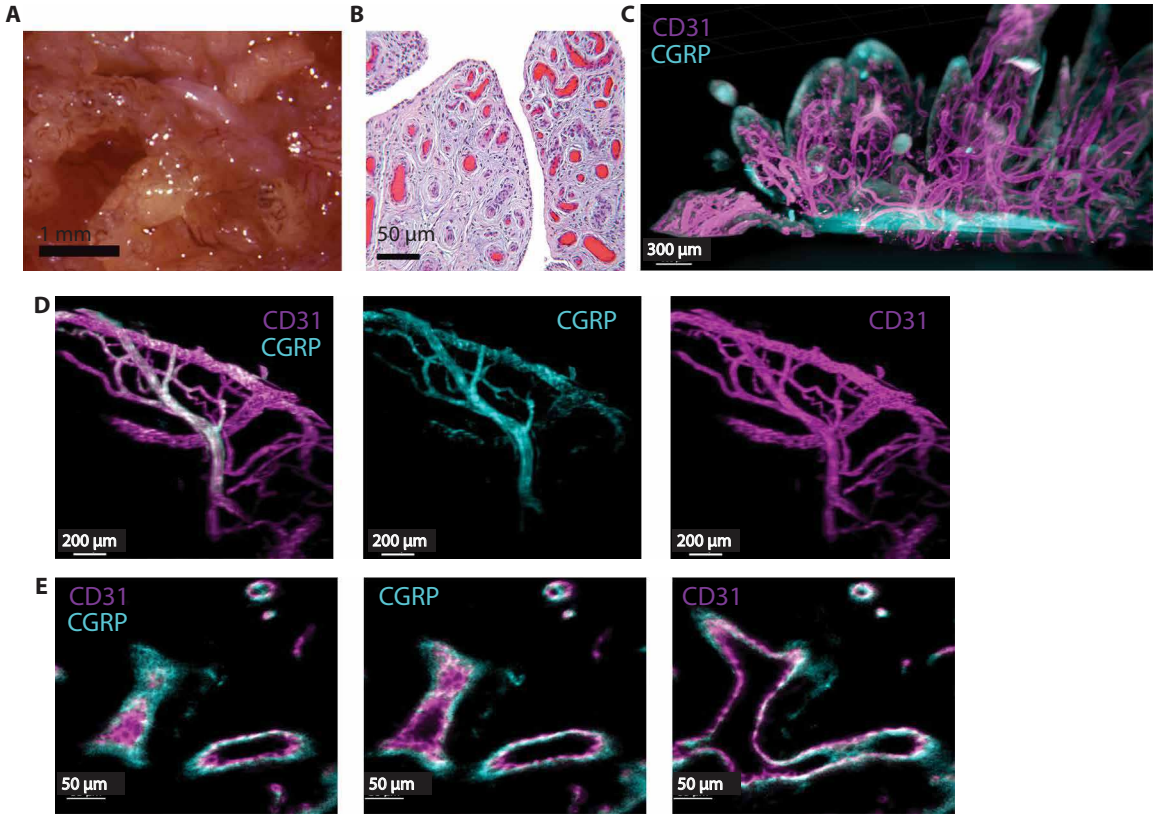

Fig. 7. RA synovial papillary hypertrophy processes contain CGRP$^+$ nociceptive neurites encasing CD31$^+$ vessels. (**A**) Gross macroscopic image of RA synovial papillary hypertrophy with visible blood vessels is shown. (**B**) H&E-stained section of RA synovial papillary hypertrophy is shown with minimal lymphocytic infiltration. Scale bar, 50 µm. (**C**) Whole-mount imaging (iDISCO) is shown from low inflammatory RA synovium stained with anti-CD31 antibodies (magenta) and anti-CGRP antibodies (cyan) (see also movie S1). Scale bar, 300 µm. (**D**) Whole-mount imaging is shown from inside a papillary process stained with anti-CD31 antibodies (magenta) and anti-CGRP antibodies (cyan) (see also movie S2). Scale bars, 200 µm. (**E**) Three consecutive optical cross sections of synovial tissue stained with anti-CD31 antibodies (magenta) and anti-CGRP antibodies (cyan) are shown. Scale bars, 50 µm.

genes in an unsupervised manner, on the basis of which genes tend to be expressed together in the dataset. In contrast, our GbGMI approach groups genes on the basis of their association with the variable of interest and was developed specifically to mitigate the challenge of relating patient-reported outcome data, which is notoriously variable, to tissue transcriptomic data, which is often inadequately powered for statistical analysis. Our approach might be useful in other scenarios where investigators are searching for clinical associations in a complex dataset with a limited sample size.

As shown in our statistical analysis, each individual gene expression did not correlate with pain. On the other hand, the coordinated expression within the GbGMI-identified pain-associated gene module suggests functional relatedness. Future studies are needed to further unravel how genes identified here relate to neuron growth and patient experience of pain either independently or in conjunction with other genes. However, our ligand-receptor interaction analysis strongly suggests that changes in gene expression within the joint drive interactions with nociceptors to promote pain. We identified numerous ligands that bind to receptors on hDRG neurons to mediate neurite/axon growth including netrins, ephrins, integrins, and semaphorins. These receptors signal through pathways known to produce pain, including phosphatidylinositol 3-kinase (PI3K)/AKT signaling (*43*). We also identified possible cross-talk between ligands expressed by hDRGs that bind to receptors encoded in the 815 GbGMI-identified pain-associated genes expressed by fibroblasts, suggesting that neurons may play a key role in influencing fibroblast differentiation through neuronal-released factors including transforming growth factor–β, fibroblast growth factors, and chemokines.

Our study has several limitations. Pain is a multidimensional experience and was measured here only by patient report, using two different questionnaires (HOOS/KOOS and VAS), and the tissue studied was limited to the synovium. Given the small sample size, other factors known to contribute to experience of RA pain, such as sex, age, disease activity, as well as central sensitization and lower overall pain tolerance, were not explored here. In addition, the experiments on neuron growth in vitro tested the effects of human protein on mouse neurons. Future studies are needed to explore the effects of fibroblast products on human neurons. There are also many other genes associated with patient report of pain in this dataset that warrant additional study. These factors include heparin binding epidermal growth factor (HBEGF), betacellulin (BTC), and connective tissue growth factor (CTGF), all of which have been shown to augment neuronal sprouting in response to injury and/or have been associated with pain in other clinical conditions or animal models (*44*–*46*). In addition, CGRP$^+$ afferents perform other

functions in the periphery besides sensing and relaying pain centrally, such as altering blood flow (*47*). Similarly, neural guidance molecules also have important functions outside the nervous system, such as affecting angiogenesis, lung branching morphogenesis, immunomodulation, and tumorigenesis, and these nonneuronal effects of this family of molecules warrant further study in the context of arthritis. Last, fluorescence of the synovial lining layer cannot be discerned from technical edge artifacts in the tissue-cleared images.

Together, this work identifies a group of genes associated with patient report of pain in low inflammatory synovium in both early untreated and established RA. These findings have the potential to be leveraged to develop next-generation therapeutic approaches to alleviating pain in RA, particularly in the low inflammatory pathotype that may be less responsive to current therapeutics that target adaptive immune inflammation.

MATERIALS AND METHODS

Study design

This study involved 139 patients with RA undergoing arthroplasty at the Hospital for Special Surgery (HSS) in New York, meeting either the American College of Rheumatology (ACR)/European League Against Rheumatism 2010 Classification criteria (*48*) and/or the ACR 1987 criteria for RA (*49*). Patient data, including KOOS and HOOS questionnaires, were collected with approval from relevant institutional review boards (detailed in the "Study approval" section) with signed informed consent from all participants. RNA was extracted from 39 bulk synovial tissue samples previously sequenced (*9*) (ImmPort accession #SDY1299), which included three gene expression clusters correlating with varying degrees of synovial inflammation: low ($n = 14$), intermediate ($n = 11$), and high ($n = 14$) (*9*). Among these 39 patients, 38 had mean nucleus density data (*16*); of these, 26 had low inflammatory synovium, among which 22 had HOOS/KOOS pain scores. External validation of pain-associated genes used bulk RNA-seq data from 87 patients with early RA obtained from the PEAC. The study aimed to uncover the source of joint pain in RA because it does not always correlate with synovial inflammation. We developed a graph-based machine-learning approach to identify a set of pain-associated genes in patients with RA and limited synovial inflammation. Further analyses were conducted to reveal which cell subtype in synovium primarily expressed the identified pain-associated genes. Studies through in vitro and in vivo approaches were conducted to test whether and how the identified pain-associated genes can suggest a potential mechanism for enhancing nociceptor growth into synovial pannus, thereby contributing to RA-associated pain. No sample size calculation, randomization, or blinding was performed.

Study approval

This study includes data from 139 patients with RA undergoing arthroplasty at the HSS in New York. All patients met either the ACR/European League Against Rheumatism 2010 Classification criteria (*48*) and/or the ACR 1987 criteria for RA (*49*). Patient demographic data and RA disease activity scores and treatments are listed in data file S9. Patient data including KOOS (*50*) and HOOS (*51*) questionnaires were collected. RA pain scores indicate response to the question, "how much pain have you felt due to your rheumatoid arthritis during the last week?" with responses ranging from 1 to 10. Condition pain scores indicate response to the question, "how much pain have you had because of your condition over the past week? Please indicate how severe your pain has been," with responses ranging from 1 to 10. This study was approved by the HSS Institutional Review Board (approval no. 2014-233), the Rockefeller University Institutional Review Board (approval no. DOR0822), and the Biomedical Research Alliance of New York (approval no. 15-08-114-385). All participating patients provided their signed informed consent.

H&E histologic scoring

Synovial samples were obtained from the most grossly inflamed (dull and opaque) area of synovium. If there were no obviously inflamed areas, then samples were obtained from standard locations: the femoral aspects of the medial and lateral gutters and the central supratrochlear region of the suprapatellar pouch. Each tissue biopsy was sectioned at 5-μm thickness and stained with Harris modified hematoxylin solution and eosin Y (H&E) manufactured by Epredia. An expert musculoskeletal pathologist scored 14 synovial histologic features in a single section for each patient: lymphocytic inflammation, mucoid change, fibrosis, fibrin, germinal centers, lining hyperplasia, neutrophils, detritus, plasma cells, binucleated plasma cells, Russell bodies, sublining giant cells, synovial lining giant cells, and mast cells. Detailed methods for scoring these features are described in prior studies (*9*) and available at www.hss.edu/pathology-synovitis, and the classification algorithm is available at ImmPort https://immport.org/shared/study/SDY1299. Spearman correlation was used to compare the relationship of patient reports of pain, which was not normally distributed, to cell density.

Gene expression established arthritis cohort

RNA was extracted from 39 bulk synovial tissue samples collected from an established RA cohort and previously sequenced as described in (*9*) (ImmPort accession #SDY1299). Briefly, these libraries were prepared using TruSeq mRNA Stranded Library kits, 50-bp paired-end reads were sequenced on a HiSeq 2500 platform, and reads were aligned to hg19 using STAR (*52*). Samples with >0.1% globin mRNA were excluded from further analysis. After quality control, ComBaT in the Bioconductor SVA package (*53*) was used for batch effect correction, and DESeq2 (*54*) was used to normalize the data. Consensus clustering identified three gene expression clusters characterized by different amounts of synovial inflammation: low ($n = 14$), intermediate ($n = 11$), and high ($n = 14$) (*9*). Of these 39 patients with RA, 38 had mean nucleus density data for the benchmarking analysis (*16*), and 26 had low inflammatory synovium, for which all 26 had scores for lymphocytic inflammation and lining hyperplasia and 22 had scores of fibrosis and HOOS/KOOS pain scores. We used limma (*15*) to test for gene expression correlates of fibrosis, lymphocytic inflammation, lining hyperplasia, and pain.

Gene expression early arthritis cohort

For external validation of the pain-associated genes identified by our method, we downloaded the bulk RNA-seq of patients with early RA from the PEAC on ArrayExpress (www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-6141/samples/). Using only samples with sufficient data quality, reads were pseudoaligned to hg38 with Kallisto to generate a counts matrix. Counts were normalized for read depth [counts into $\log_2(\text{CPM} + 1)$ (counts per million)] and batch-corrected using removeBatchEffect from limma (*15*).

Gene expression in bulk sorted cells and single cells

In the analyses related to RA synovial cell types, we used the sorted-population bulk RNA-seq gene expression data of the synovial T cells ($CD45^+$, $CD3^+$, and $CD14^-$), monocytes ($CD45^+$, $CD3^-$, and $CD14^+$), B cells ($CD45^+$, $CD3^-$, $CD14^-$, and $CD19^+$), and fibroblasts ($CD45^-$, $CD31^-$, and $PDPN^+$) collected by fluorescence-activated cell sorting (BD FACSAria Fusion) directly in buffer RLT (QIAGEN)(*19*). Normalized read counts per gene as transcripts per million (TPMs) were used (ImmPort accession #SDY998) (*10*).

We used the scRNA-seq data preprocessed in (*19*), where the gene expression amounts were quantified by counting UMIs (unique molecular identifiers) and transformed into $\log_2(\mathrm{CPM} + 1)$. From the scRNA-seq profiles for 32,391 genes and 5265 cells left after rigorous quality control, 18 unique cell populations were identified by an integrated strategy based on canonical correlation analysis. Specifically, in fibroblasts: $CD34^+$ sublining fibroblasts (SC-F1), human leukocyte antigen (HLA)–DRA^{hi} sublining fibroblasts (SC-F2), Dickkopf-related protein $(DKK3)^+$ sublining fibroblasts (SC-F3), and $CD55^+$ lining fibroblasts (SC-F4); in monocytes: interleukin-1B $(IL\text{-}1B)^+$ proinflammatory monocytes (SC-M1), nuclear protein 1 $(NUPR1)^+$ monocytes (SC-M2), $C1QA^+$ monocytes (SC-M3), and interferon-activated monocytes (SC-M4); in T cells: three $CD4^+$ clusters: $CCR7^+$ T cells (SC-T1), $FOXP3^+$ regulatory T cells (SC-T2), and programmed cell death protein 1 $(PDCD1)^+$ T peripheral helper (T_{PH}) and T follicular helper (T_{FH}) (SC-T3) cells, and three $CD8^+$ clusters, $GZMK^+$ T cells (SC-T4), $GNLY^+GZMB^+$ cytotoxic lymphocytes (SC-T5), and $GZMK^+GZMB^+$ T cells (SC-T6); in B cells: naïve $IGHD^+CD27^-$ (SC-B1), $IGHG3^+$ $CD27^+$ memory B cells (SC-B2), autoimmune-associated B cell cluster (SC-B3) with high expression of integrin alpha-X (ITGAX) (also known as CD11c), and a plasmablast cluster (SC-B4) with high expression of immunoglobulin genes and X-box-binding protein 1 (XBP1) (*19*). The top 200 marker genes for each subpopulation identified by differential expression analysis in the original publication were used as signature genes for individual cell types within RA synovium.

Description of our GbGMI framework

We developed GbGMI, a graph-based machine-learning framework of algorithms, to identify a gene expression module that strongly correlates with pain in low inflammatory RA synovial tissues (fig. S1). We use $M \in R^{m \times n}$ to represent the input gene expression matrix of genes and patients. We use a numeric vector $\boldsymbol{a} \in R^n$ to represent the pain scores reported by these patients. We quantified the quality Q of a selected gene subset through the correlation between their collective expression and the pain score and then searched for a gene subset with optimal quality Q as a feature selection task (*55*–*58*). Because an exhaustive search through all possible subsets of an input gene set to optimize Q is computationally intractable (*59*), we adapted and integrated the feature scoring strategy used in the filter feature-selection/ranking approaches (*60*–*62*) and the feature subset scoring strategy used in the wrapper feature-selection approaches (*58, 63*). Specifically, we generated a gene prioritization list L_s according to how well each individual gene expression respects the geometric structure over patients built according to their pain scores. Then, the quality Q_k of the kth candidate gene subset comprising the top k genes in L_s was evaluated for k ranging from 1 to m. The first k^* where Q_k peaked was used as the cutoff point on L_s. This subset of k^* genes is the output pain-associated gene module.

We next adapted Laplacian score algorithm to prioritize genes. Given the input gene expression matrix and pain-score vector, we ranked the genes by the way each gene expression vector (that is, a row vector in the gene expression matrix M) respected a given geometric structure over the patients encoded in an $n \times n$ similarity matrix S based on the pain score vector $\boldsymbol{a}$, by adapting the Laplacian score algorithm (*17*), instead of assuming independent observations in the correlation tests (fig. S1, step A). To compute S, a Gaussian kernel, which empirically outperformed other types of kernels (*64, 65*), was adopted to map the Euclidean distance between the pain scores of each pair of patients and into a similarity measure $S(i, j) \in [0,1]$

$$S(i,j) = e - \frac{|\,\boldsymbol{a}(i) - \boldsymbol{a}(j)\,|^2}{h} \tag{1}$$

where h corresponds to the bandwidth or smoothing factor in a kernel metric definition. This formulation forces the similarity measure between any two patients with significantly different pain scores to be close to 0 while pushing the similarity measure between the patients of pain scores within a certain range (depending on the smoothing factor) to be closer to 1. This promotes the locality we want to focus on. We set h to control the local neighborhood of patients on graph S according to the theoretical range of the pain scores (for example, $h = 100$ for HSS HOOS/KOOS pain score or PEAC VAS characteristic because either ranges between 0 and 100).

The Laplacian score of each gene was then computed to evaluate how well this gene's expression on these patients preserves S (fig. S1, step B). This is different from the original publication (*17, 66*) that aimed to preserve the input feature space (for example, the input low inflammatory genes), wherein we aimed to select features (for example, the k^* pain-associated genes). The Laplacian matrix of S is defined by $L = D - S$, where D is a diagonal matrix with D_{ii} indicating the degree of node (that is, patient) i in the weighted graph S [that is, $D = \mathrm{diag}(S\,\mathbf{1})$]. For the rth gene, let $M_{r,*}$ be its n-dimensional gene expression vector across the patients (that is, the rth row vector in the gene expression matrix M), its Laplacian score $\mathrm{ls}(r)$ is computed as

$$ls(r) = \frac{\widehat{\boldsymbol{m}_r}^T L\, \widehat{\boldsymbol{m}_r}}{\widehat{\boldsymbol{m}_r}^T D\, \widehat{\boldsymbol{m}_r}}, \widehat{\boldsymbol{m}_r} = {M_{r,*}}^T - \frac{M_{r,*} D\mathbf{1}}{\mathbf{1}^T D\mathbf{1}}\mathbf{1} \tag{2}$$

where the symbol $\mathbf{1}$ denotes a column vector whose all elements are 1's with dimensionality determined by context. The gene prioritization list L_s is generated by sorting the input m genes according to their Laplacian scores in ascending order. The smaller the Laplacian score of a gene, the better its expression data respect the geometric structure defined by S over the patients [according to objective function analysis (*17*)]. Each top k subset on L_s forms a candidate gene subset for selection. The rows of M were reordered according to L_s (fig. S1, step C).

Quantification of candidate k-gene subset quality Q_k

The Q_k of the kth candidate gene subset was measured on the basis of the association between their collective expression pattern and the pain-score vector $\boldsymbol{a}$. Given the gene expression submatrix $M_{1:k,*}$ of this candidate k-gene subset, Q_k was computed through two steps: First, project the k-gene expression vector of each patient into a univariate summary score that preserves the patient-to-patient similarity structure in the original k-dimensional feature (gene)

space. This addresses the dimensionality mismatch between the multigene expression and the univariate patient-level pain score (*9*, *17*). The resulting summary score vector of the n patients is denoted $\boldsymbol{s}_k$ (fig. S1, step D). Second, quantify Q_k using the statistical significance of correlation test [for example, the − log (P value) of Kendall's correlation test] between $\boldsymbol{s}_k$ and $\boldsymbol{a}$ over the same patients. We chose the first k^*, where $Q_k{}^*$ peaked, as the cutoff point on the sorted gene prioritization list L_s. This subset of k^* genes is the pain-associated gene module identified by our GbGMI framework (fig. S1, step E).

For fig. S1 (step D), we used the t-SNE (*66*). The resulting summary score vector $\boldsymbol{s}_k$ for the n patients respects how the gene expression data were arranged in the k-gene feature space of this candidate gene subset. We hereon briefly sketch the instantiation of t-SNE with the variables involved in our study. Following the SNE framework (*67*), the directional similarity of patient j to patient i based on their multigene expression vectors $M_{1:k,i}$ and $M_{1:k,j}$ is

$$p(j\,|\,i) = \frac{\exp(-\,\|\,M_{1:k,i} - M_{1:k,j}\,\|^2 / 2\sigma_i{}^2)}{\Sigma_{l\neq i}\exp(-\,\|\,M_{1:k,i} - M_{1:k,l}\,\|^2 / 2\sigma_i{}^2)} \tag{3}$$

where the variance of the Gaussian kernel $\sigma_i{}^2$ is chosen such that the perplexity of the conditional probability distribution over all points $j \neq i$ defined by

$$\mathrm{Perp}(P_i) = 2H(P_i), H(P_i) = -\,\Sigma_j\, p(j\,|\,i)\,\log_2 p(j\,|\,i) \tag{4}$$

matches a prespecified value. The perplexity in this context can be interpreted as an estimation about the number of close neighbors of each patient on graph S. Therefore, we specified the perplexity on the basis of the rounded mean degree of the similarity graph S built from $\boldsymbol{a}$. The symmetric SNE was used for mathematical and computational convenience in the t-SNE formulation by defining the following undirected similarities

$$p(i,j) = \frac{p(j\,|\,i) + p(i\,|\,j)}{2n} \tag{5}$$

where n is the number of patients. Because $\Sigma_{i,j}p(i,j) = 1$, this is a valid probability distribution on the set of all pairs (i,j). The t-SNE step in this algorithm uses the t-distribution with one degree of freedom (also known as Cauchy distribution) as the one-dimensional similarity kernel applied to pairs of summary scores defined by

$$q(i,j) = \frac{(1 + \|\boldsymbol{s}(i) - \boldsymbol{s}(j)\|^2)^{-1}}{\Sigma_{k\neq l}(1 + \|\boldsymbol{s}(k) - \boldsymbol{s}(l)\|^2)^{-1}} \tag{6}$$

The main idea of this t-SNE-based step is to arrange the patients in a one-dimensional space such that the similarities $q(i,j)$ between $\boldsymbol{s}(i)$ and $\boldsymbol{s}(j)$ match $p(i,j)$ as close as possible in terms of the Kullback-Leibler divergence. Thus, the loss function is

$$L = \Sigma_{i,j}p(i,j)\log\frac{p(i,j)}{q(i,j)} \tag{7}$$

a summary score vector $\boldsymbol{s}_k$ is computed for each candidate k-gene subset, rendering a set of m such vectors $\{\boldsymbol{s}_k \mid k = 1,2,\cdots,m\}$.

Our framework GbGMI can be applied to identify a subset of genes that collectively associate with some other patient-level numeric attribute beyond the main focus of this paper (the HOOS/KOOS pain score). Our framework is also flexible and adaptive in different contexts by replacing specific computational components with other design choices (for example, other embeddings instead of t-SNE for computing summary scores of selected gene expression).

DRG dissection and digestion

Before extraction of the DRG neurons, chambered coverslips (ibidi, 80286) were coated overnight at 37°C with poly-L-lysine (Sigma-Aldrich, P4832) and then for 2 hours at 37°C with mouse sarcoma basement membrane laminin (Sigma-Aldrich, L2020) diluted 1:50 with 1× phosphate-buffered saline (PBS) and with one also coated with huNTN4 (0.2 μg/ml; R&D Systems, 1254-N4). For each assay, sensory DRG neurons were harvested from two female 6- to 8-week-old C57BL/6 mice (the Jackson Laboratory), under a dissection microscope using forceps, and placed into a waiting 15-ml conical tube on ice containing 1× L-15 medium (Thermo Fisher Scientific, 21083027). DRGs were spun down at 950 rpm for 2 min. Medium was aspirated and replaced with 1 ml of L-15 containing dispase II (10 mg/ml; Sigma-Aldrich, 04942078001) and collagenase IV (10 mg/ml; Thermo Fisher Scientific, 17104019). DRGs were then placed at 37°C for 20 min. Enzyme solution was then carefully aspirated and replaced with 2 ml of L-15. Pellet was resuspended thoroughly with a 1000-μl pipette. Twenty-five microliters of deoxyribonuclease I (10 mg/ml) was then added, and once again, the cells were placed at 37°C for 20 min. The cells were then spun down for 5 min at 950 rpm and resuspended in 5 ml of L-15. After another 5 min of centrifugation, the cells were resuspended in 1 ml of L-15 and layered on top of 15% ice-cold bovine serum albumin (Sigma-Aldrich, A7906) and spun down at room temperature for 8 min at 1179 rpm to remove myelin. The cell pellet was resuspended in 1× Neurobasal Plus medium (Thermo Fisher Scientific, A3582901) containing B27 (Thermo Fisher Scientific, 17504001) diluted 1:50, GlutaMAX (Thermo Fisher Scientific, 35050061) diluted 1:100, and gentamicin sulfate (Abbott Laboratories) and then plated onto the precoated slides. Human β-NGF (0.1 μg/ml; R&D Systems, 256-GF) was added to the medium of the positive control slide chamber.

Harvesting fibroblast supernatant

A total of 500,000 synovial fibroblasts from low inflammatory RA, between five and eight passages, were plated in each well of six-well cell culture plates and covered with 2 ml of Neurobasal medium (Thermo Fisher Scientific, 21103049) with added 2% B-27 supplement (Thermo Fisher Scientific, 17504044), 1% GlutaMAX supplement (Thermo Fisher Scientific, 35050061), and 0.04% gentamicin sulfate. Medium was changed, and cells were washed in PBS after 24 hours at 37°C to allow for recovery from freeze-thaw. Fresh medium was added, and cells were left at 37°C overnight. After 24 hours, medium was removed and centrifuged at 1300 rpm for 5 min to remove cells and cell debris. Supernatant was then collected and stored at −80°C.

DRG neuron culture

A total of 50,000 neurons harvested from mice were plated on poly-L-lysine and 2% laminin (Sigma-Aldrich, L2020) coated two-well chamber slides with either medium alone, fibroblast supernatant + medium (1:1), huNGF, huNTN4, or mouse Ntn4. To remove precipitates formed during freeze-thaw, medium and supernatant samples were first filtered using 50-ml MilliporeSigma Steriflip sterile disposable vacuum filters with 0.22-μm pore membranes. Neurons were cultured for 24 hours at 37°C before fixation in 4% paraformaldehyde.

Neurite imaging, quantification, and comparison

Each chambered slide was examined under a Keyence fluorescence microscope, and the accompanying software was used to capture a stitched image of a large 5 × 5 area randomly selected on the plate at ×10 magnification. Five of these images were captured per plate. To quantify sprouting versus nonsprouting DRG neurons, each image was examined in ImageJ and used the Cell Counter plugin to keep track of the total number of sprouting neurons, identified as having at least three neurites extended from the cell body where the extensions were more than twice the diameter of the cell body and exhibited some degree of branching. From across these images, $n = 10$ neurons that exhibited branching were selected, and their branching was quantified using the Sholl Analysis plugin. Each analysis used a start radius of 30 pixels, an end radius of 225 pixels, and a step size of 7. Mixed-model repeated-measures analysis was used to analyze Sholl data. The model included group, radius (categorical), and group*radius interaction as fixed effects. A significant group*radius interaction indicated group differences and branching.

iDISCO/imaging

Synovial tissue obtained was fixed overnight in 4% paraformaldehyde and washed three times with PBS for 5 min. Samples were stored long-term in PBS with 0.03% sodium azide and were stained following iDISCO protocol. Briefly, samples were treated as indicated using alternative nonmethanol pretreatment protocol before immunolabeling. All steps assumed an $n = 1$. Antibodies used for staining were used at 1:150 and include anti-CD31 antibody polyclonal guinea pig (Synaptic Systems, 351004), anti-CGRP antibody polyclonal goat (ab36001, Abcam), Alexa Fluor 488 donkey anti-mouse immunoglobulin G (IgG) (H + L) ReadyProbes (Invitrogen), and Alexa Fluor 568 donkey F(ab′)2 anti-rabbit IgG (H + L; Abcam). After immunolabeling, samples were embedded in agar before clearing with dibenzyl ether (DBE). Samples were imaged by light sheet microscopy, submerged in DBE on an Ultramicroscope II (LaVision/Miltenyi BioTec). Images were captured using a 4× objective with 2.5-μm Z-slices. Acquired images were visualized using Imaris ×64 software (version 9.1 Bitplane), and three-dimensional reconstructions were recorded as mp4 video files. Optical slices were generated using the orthoslicer or oblique slicer tool.

Statistical analysis

Statistical analysis was performed with functions or libraries from the R software package unless stated otherwise. A P value, raw or adjusted depending on the context, below 0.05 was considered significant. For comparison between pain scores in two patient groups, data were analyzed by unpaired, two-sided Mann-Whitney. A moderated t test from limma (*15*) was used for testing individual gene expression correlates of fibrosis, lymphocytic inflammation, lining hyperplasia, and pain. To identify differentially expressed genes, a one-way analysis of variance (ANOVA) test was performed to compare individual gene expression among different pain level–based groupings. The FDR was used in multiple hypothesis testing to correct for multiple comparisons where applicable. For assessing the correlation between two patient-level scores (of the three: gene expression summary score, pain score, or cell density), Spearman's or Kendall's rank correlation tests were used. For comparison between group level pain–associated gene expression and synovial tissue gene expression, Kolmogorov-Smirnov test was used. To compare the quantification among different groups of DRG neurons, data were analyzed by unpaired, two-sided Kruskal-Wallis test or two-way ANOVA test, followed by Dunn's posttest ($*P < 0.05$, $**P < 0.01$, $***P < 0.001$, and $****P < 0.0001$). A hypergeometric test was performed to estimate statistical significance in pathway or gene set enrichment analysis, adjusted for multiple testing using g:SCS correction (*68*).

Supplementary Materials

This PDF file includes:
Materials and Methods
Figs. S1 to S9
References (*69–83*)

Other Supplementary Material for this manuscript includes the following:
Data files S1 to S9
Movies S1 and S2
MDAR Reproducibility Checklist

REFERENCES AND NOTES

1. F. Marchand, M. Perretti, S. B. McMahon, Role of the immune system in chronic pain. *Nat. Rev. Neurosci.* **6**, 521–532 (2005).
2. S. Alivernini, G. S. Firestein, I. B. McInnes, The pathogenesis of rheumatoid arthritis. *Immunity* **55**, 2255–2270 (2022).
3. M. H. Buch, S. Eyre, D. McGonagle, Persistent inflammatory and non-inflammatory mechanisms in refractory rheumatoid arthritis. *Nat. Rev. Rheumatol.* **17**, 17–33 (2021).
4. G. Nagy, N. M. Roodenrijs, P. M. Welsing, M. Kedves, A. Hamar, M. C. van der Goes, A. Kent, M. Bakkers, E. Blaas, L. Senolt, Z. Szekanecz, E. Choy, M. Dougados, J. W. Jacobs, R. Geenen, H. W. Bijlsma, A. Zink, D. Aletaha, L. Schoneveld, P. van Riel, L. Gutermann, Y. Prior, E. Nikiphorou, G. Ferraccioli, G. Schett, K. L. Hyrich, U. Mueller-Ladner, M. H. Buch, I. B. McInnes, D. van der Heijde, J. M. van Laar, EULAR definition of difficult-to-treat rheumatoid arthritis. *Ann. Rheum. Dis.* **80**, 31–35 (2021).
5. Y. C. Lee, M. L. Frits, C. K. Iannaccone, M. E. Weinblatt, N. A. Shadick, D. A. Williams, J. Cui, Subgrouping of patients with rheumatoid arthritis based on pain, fatigue, inflammation, and psychosocial factors. *Arthritis Rheum.* **66**, 2006–2014 (2014).
6. D. F. McWilliams, S. Rahman, R. J. E. James, E. Ferguson, P. D. W. Kiely, A. Young, D. A. Walsh, Disease activity flares and pain flares in an early rheumatoid arthritis inception cohort; characteristics, antecedents and sequelae. *BMC Rheumatol.* **3**, 49 (2019).
7. L. C. Pollard, E. H. Choy, J. Gonzalez, B. Khoshaba, D. L. Scott, Fatigue in rheumatoid arthritis reflects pain, not disease activity. *Rheumatology* **45**, 885–889 (2006).
8. S. Stebbings, P. Herbison, T. C. H. Doyle, G. J. Treharne, J. Highton, A comparison of fatigue correlates in rheumatoid arthritis and osteoarthritis: Disparity in associations with disability, anxiety and sleep disturbance. *Rheumatology* **49**, 361–367 (2010).
9. D. E. Orange, P. Agius, E. F. DiCarlo, N. Robine, H. Geiger, J. Szymonifka, M. McNamara, R. Cummings, K. M. Andersen, S. Mirza, M. Figgie, L. B. Ivashkiv, A. B. Pernis, C. S. Jiang, M. O. Frank, R. B. Darnell, N. Lingampali, W. H. Robinson, E. Gravallese, Accelerating Medicines Partnership in Rheumatoid Arthritis and Lupus Network, V. P. Bykerk, S. M. Goodman, L. T. Donlin, Identification of three rheumatoid arthritis disease subtypes by machine learning integration of synovial histologic features and RNA sequencing data. *Arthritis Rheum.* **70**, 690–701 (2018).
10. G. Dennis, C. T. Holweg, S. K. Kummerfeld, D. F. Choy, A. F. Setiadi, J. A. Hackney, P. M. Haverty, H. Gilbert, W. Y. Lin, L. Diehl, S. Fischer, A. Song, D. Musselman, M. Klearman, C. Gabay, A. Kavanaugh, J. Endres, D. A. Fox, F. Martin, M. J. Townsend, Synovial phenotypes in rheumatoid arthritis correlate with response to biologic therapeutics. *Arthritis Res. Ther.* **16**, ar4555 (2014).
11. M. J. Lewis, M. R. Barnes, K. Blighe, K. Goldmann, S. Rana, J. A. Hackney, N. Ramamoorthi, C. R. John, D. S. Watson, S. K. Kummerfeld, R. Hands, S. Riahi, V. Rocher-Ros, F. Rivellese, F. Humby, S. Kelly, M. Bombardieri, N. Ng, M. DiCicco, D. van der Heijde, R. Landewé, A. van der Helm-van Mil, A. Cauli, I. B. McInnes, C. D. Buckley, E. Choy, P. C. Taylor, M. J. Townsend, C. Pitzalis, Molecular portraits of early rheumatoid arthritis identify clinical and treatment response phenotypes. *Cell Rep.* **28**, 2455–2470.e5 (2019).
12. F. Humby, M. Lewis, N. Ramamoorthi, J. A. Hackney, M. R. Barnes, M. Bombardieri, A. F. Setiadi, S. Kelly, F. Bene, M. DiCicco, S. Riahi, V. Rocher, N. Ng, I. Lazarou, R. Hands, D. van der Heijde, R. B. M. Landewé, A. van der Helm-van Mil, A. Cauli, I. McInnes, C. D. Buckley, E. H. Choy, P. C. Taylor, M. J. Townsend, C. Pitzalis, Synovial cellular and molecular signatures stratify clinical response to csDMARD therapy and predict radiographic progression in early rheumatoid arthritis patients. *Ann. Rheum. Dis.* **78**, 761–772 (2019).

13. F. Zhang, A. H. Jonsson, A. Nathan, N. Millard, M. Curtis, Q. Xiao, M. Gutierrez-Arcelus, W. Apruzzese, G. F. M. Watts, D. Weisenfeld, S. Nayar, J. Rangel-Moreno, N. Meednu, K. E. Marks, I. Mantel, J. B. Kang, L. Rumker, J. Mears, K. Slowikowski, K. Weinand, D. E. Orange, L. Geraldino-Pardilla, K. D. Deane, D. Tabechian, A. Ceponis, G. S. Firestein, M. Maybury, I. Sahbudin, A. Ben-Artzi, A. M. Mandelin, A. Nerviani, M. J. Lewis, F. Rivellese, C. Pitzalis, L. B. Hughes, D. Horowitz, E. DiCarlo, E. M. Gravallese, B. F. Boyce, L. W. Moreland, S. M. Goodman, H. Perlman, V. M. Holers, K. P. Liao, A. Filer, V. P. Bykerk, K. Wei, D. A. Rao, L. T. Donlin, J. H. Anolik, M. B. Brenner, S. Raychaudhuri, Deconstruction of rheumatoid arthritis synovium defines inflammatory subtypes. *Nature* **623**, 616–624 (2023).
14. A. Nerviani, M. Di Cicco, A. Mahto, G. Lliso-Ribera, F. Rivellese, G. Thorborn, R. Hands, M. Bellan, D. Mauro, M.-A. Boutet, G. Giorli, M. Lewis, S. Kelly, M. Bombardieri, F. Humby, C. Pitzalis, A pauci-immune synovial pathotype predicts inadequate response to TNFα-blockade in rheumatoid arthritis patients. *Front. Immunol.* **11**, 845 (2020).
15. M. E. Ritchie, B. Phipson, D. Wu, Y. Hu, C. W. Law, W. Shi, G. K. Smyth, limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
16. S. Guan, B. Mehta, D. Slater, J. R. Thompson, E. DiCarlo, T. Pannellini, D. Pearce-Fisher, F. Zhang, S. Raychaudhuri, C. Hale, C. S. Jiang, S. Goodman, D. E. Orange, Rheumatoid arthritis synovial inflammation quantification using computer vision. *ACR Open Rheumatol* **4**, 322–331 (2022).
17. X. He, D. Cai, P. Niyogi, "Laplacian score for feature selection," 2005; https://proceedings.neurips.cc/paper_files/paper/2005/file/b5b03f06271f8917685d14cea7c6c50a-Paper.pdf.
18. S. Alivernini, L. MacDonald, A. Elmesmari, S. Finlay, B. Tolusso, M. R. Gigante, L. Petricca, C. Di Mario, L. Bui, S. Perniola, M. Attar, M. Gessi, A. L. Fedele, S. Chilaka, D. Somma, S. N. Sansom, A. Filer, C. McSharry, N. L. Millar, K. Kirschner, A. Nerviani, M. J. Lewis, C. Pitzalis, A. R. Clark, G. Ferraccioli, I. Udalova, C. D. Buckley, E. Gremese, I. B. McInnes, T. D. Otto, M. Kurowska-Stolarska, Distinct synovial tissue macrophage subsets regulate inflammation and remission in rheumatoid arthritis. *Nat. Med.* **26**, 1295–1306 (2020).
19. F. Zhang, K. Wei, K. Slowikowski, C. Y. Fonseka, D. A. Rao, S. Kelly, S. M. Goodman, D. Tabechian, L. B. Hughes, K. Salomon-Escoto, G. F. M. Watts, A. H. Jonsson, J. Rangel-Moreno, N. Meednu, C. Rozo, W. Apruzzese, T. M. Eisenhaure, D. J. Lieb, D. L. Boyle, A. M. Mandelin II, Accelerating Medicines Partnership Rheumatoid Arthritis and Systemic Lupus Erythematosus (AMP RA/SLE) Consortium, B. F. Boyce, E. DiCarlo, E. M. Gravallese, P. K. Gregersen, L. Moreland, G. S. Firestein, N. Hacohen, C. Nusbaum, J. A. Lederer, H. Perlman, C. Pitzalis, A. Filer, V. M. Holers, V. P. Bykerk, L. T. Donlin, J. H. Anolik, M. B. Brenner, S. Raychaudhuri, Defining inflammatory cell states in rheumatoid arthritis joint synovial tissues by integrating single-cell transcriptomics and mass cytometry. *Nat. Immunol* **20**, 928–942 (2019).
20. M. Jung, M. Dourado, J. Maksymetz, A. Jacobson, B. I. Laufer, M. Baca, O. Foreman, D. H. Hackos, L. Riol-Blanco, J. S. Kaminker, Cross-species transcriptomic atlas of dorsal root ganglia reveals species-specific programs for sensory function. *Nat. Commun.* **14**, 366 (2023).
21. S. A. Bhuiyan, M. Xu, L. Yang, E. Semizoglou, P. Bhatia, K. I. Pantaleo, I. Tochitsky, A. Jain, B. Erdogan, S. Blair, V. Cat, J. M. Mwirigi, I. Sankaranarayanan, D. Tavares-Ferreira, U. Green, L. A. McIlvried, B. A. Copits, Z. Bertels, J. S. Del Rosario, A. J. Widman, R. A. Slivicki, J. Yi, C. J. Woolf, J. K. Lennerz, J. L. Whited, T. J. Price, R. W. Gereau 4th, W. Renthal, Harmonized cross-species cell atlases of trigeminal and dorsal root ganglia. bioRxiv 2023.07.04.547740 [Preprint] (2023); https://doi.org/10.1101/2023.07.04.547740.
22. D. Tavares-Ferreira, S. Shiers, P. R. Ray, A. Wangzhou, V. Jeevakumar, I. Sankaranarayanan, A. M. Cervantes, J. C. Reese, A. Chamessian, B. A. Copits, P. M. Dougherty, R. W. Gereau IV, M. D. Burton, G. Dussor, T. J. Price, Spatial transcriptomics of dorsal root ganglia identifies molecular signatures of human nociceptors. *Sci. Transl. Med.* **14**, eabj8186 (2022).
23. R. Y. North, Y. Li, P. Ray, L. D. Rhines, C. E. Tatsui, G. Rao, C. A. Johansson, H. Zhang, Y. H. Kim, B. Zhang, G. Dussor, T. H. Kim, T. J. Price, P. M. Dougherty, Electrophysiological and transcriptomic correlates of neuropathic pain in human dorsal root ganglion neurons. *Brain* **142**, 1215–1226 (2019).
24. N. Sharma, K. Flaherty, K. Lezgiyeva, D. E. Wagner, A. M. Klein, D. D. Ginty, The emergence of transcriptional identity in somatosensory neurons. *Nature* **577**, 392–398 (2020).
25. M. V. Kuleshov, M. R. Jones, A. D. Rouillard, N. F. Fernandez, Q. Duan, Z. Wang, S. Koplev, S. L. Jenkins, K. M. Jagodnik, A. Lachmann, M. G. McDermott, C. D. Monteiro, G. W. Gundersen, A. Maayan, Enrichr: A comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44**, W90–W97 (2016).
26. E. Y. Chen, C. M. Tan, Y. Kou, Q. Duan, Z. Wang, G. V. Meirelles, N. R. Clark, A. Ma'ayan, Enrichr: Interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* **14**, 128 (2013).
27. P. I. Mapp, D. A. Walsh, N. E. Garrett, B. L. Kidd, S. C. Cruwys, J. M. Polak, D. R. Blake, Effect of three animal models of inflammation on nerve fibres in the synovium. *Ann. Rheum. Dis.* **53**, 240–246 (1994).
28. T. Serafini, T. E. Kennedy, M. J. Galko, C. Mirzayan, T. M. Jessell, M. Tessier-Lavigne, The netrins define a family of axon outgrowth-promoting proteins homologous to C. elegans UNC-6. *Cell* **78**, 409–424 (1994).
29. M. Koch, J. R. Murrell, D. D. Hunter, P. F. Olson, W. Jin, D. R. Keene, W. J. Brunken, R. E. Burgeson, A novel member of the netrin family, beta-netrin, shares homology with the beta chain of laminin: Identification, expression, and functional characterization. *J. Cell Biol.* **151**, 221–234 (2000).
30. R. Reuten, T. R. Patel, M. McDougall, N. Rama, D. Nikodemus, B. Gibert, J.-G. Delcros, C. Prein, M. Meier, S. Metzger, Z. Zhou, J. Kaltenberg, K. K. McKee, T. Bald, T. Tüting, P. Zigrino, V. Djonov, W. Bloch, H. Clausen-Schaumann, E. Poschl, P. D. Yurchenco, M. Ehrbar, P. Mehlen, J. Stetefeld, M. Koch, Structural decoding of netrin-4 reveals a regulatory function towards mature basement membranes. *Nat. Commun.* **7**, 13515 (2016).
31. Y. Hayano, K. Sasaki, N. Ohmura, M. Takemoto, Y. Maeda, T. Yamashita, Y. Hata, K. Kitada, N. Yamamoto, Netrin-4 regulates thalamocortical axon branching in an activity-dependent fashion. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 15226–15231 (2014).
32. R. M. Lindsay, Nerve growth factors (NGF, BDNF) enhance axonal regeneration but are not required for survival of adult sensory neurons. *J. Neurosci.* **8**, 2394–2405 (1988).
33. F. Denk, D. L. Bennett, S. B. McMahon, Nerve growth factor and pain mechanisms. *Annu. Rev. Neurosci.* **40**, 307–325 (2017).
34. N. Renier, Z. Wu, D. J. Simon, J. Yang, P. Ariel, M. Tessier-Lavigne, iDISCO: A simple, rapid method to immunolabel large tissue samples for volume imaging. *Cell* **159**, 896–910 (2014).
35. F. Mizoguchi, K. Slowikowski, K. Wei, J. L. Marshall, D. A. Rao, S. K. Chang, H. N. Nguyen, E. H. Noss, J. D. Turner, B. E. Earp, P. E. Blazar, J. Wright, B. P. Simmons, L. T. Donlin, G. D. Kalliolias, S. M. Goodman, V. P. Bykerk, L. B. Ivashkiv, J. A. Lederer, N. Hacohen, P. A. Nigrovic, A. Filer, C. D. Buckley, S. Raychaudhuri, M. B. Brenner, Functionally distinct disease-associated fibroblast subsets in rheumatoid arthritis. *Nat. Commun.* **9**, 789 (2018).
36. T. Iijima, J.-Q. Zhang, Three-dimensional wall structure and the innervation of dental pulp blood vessels. *Microsc. Res. Tech.* **56**, 32–41 (2002).
37. G. Gonçalves Dos Santos, J. M. Jimenéz-Andrade, S. A. Woller, E. Muñoz-Islas, M. B. Ramírez-Rosas, N. Ohashi, G. Ferreira Catroli, Y. Fujita, T. L. Yaksh, M. Corr, The neuropathic phenotype of the K/BxN transgenic mouse with spontaneous arthritis: Pain, nerve sprouting and joint remodeling. *Sci. Rep.* **10**, 15596 (2020).
38. A. M. Obeidat, R. E. Miller, R. J. Miller, A.-M. Malfait, The nociceptive innervation of the normal and osteoarthritic mouse knee. *Osteoarthr. Cartil* **27**, 1669–1679 (2019).
39. K. Aso, S. M. Shahtaheri, R. Hill, D. Wilson, D. F. McWilliams, L. N. Nwosu, V. Chapman, D. A. Walsh, Contribution of nerves within osteochondral channels to osteoarthritis knee pain in humans and rats. *Osteoarthr. Cartil* **28**, 1245–1254 (2020).
40. S. Zhu, J. Zhu, G. Zhen, Y. Hu, S. An, Y. Li, Q. Zheng, Z. Chen, Y. Yang, M. Wan, R. L. Skolasky, Y. Cao, T. Wu, B. Gao, M. Yang, M. Gao, J. Kuliwaba, S. Ni, L. Wang, C. Wu, D. Findlay, H. K. Eltzschig, H. W. Ouyang, J. Crane, F.-Q. Zhou, Y. Guan, X. Dong, X. Cao, Subchondral bone osteoclasts induce sensory innervation and osteoarthritis pain. *J. Clin. Invest.* **129**, 1076–1093 (2019).
41. D. E. Nanus, A. Badoume, S. N. Wijesinghe, A. M. Halsey, P. Hurley, Z. Ahmed, R. Botchu, E. T. Davis, M. A. Lindsay, S. W. Jones, Synovial tissue from sites of joint pain in knee osteoarthritis patients exhibits a differential phenotype with distinct fibroblast subsets. *EBioMedicine* **72**, 103618 (2021).
42. E. Lejmi, L. Leconte, S. Pédron-Mazoyer, S. Ropert, W. Raoul, S. Lavalette, I. Bouras, J.-G. Feron, M. Maitre-Boube, F. Assayag, C. Feumi, M. Alemany, T. X. Jie, T. Merkulova, M.-F. Poupon, M.-M. Ruchoux, G. Tobelem, F. Sennlaub, J. Plouët, Netrin-4 inhibits angiogenesis via binding to neogenin and recruitment of Unc5B. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 12491–12496 (2008).
43. S.-P. Chen, Y.-Q. Zhou, D.-Q. Liu, W. Zhang, A. Manyande, X.-H. Guan, Y.-K. Tian, D.-W. Ye, D. M. Omar, PI3K/Akt pathway: A potential therapeutic target for chronic pain. *Curr. Pharm. Des.* **23**, 1860–1868 (2017).
44. J. P. Borges, K. Mekhail, G. D. Fairn, C. N. Antonescu, B. E. Steinberg, Modulation of pathological pain by epidermal growth factor receptor. *Front. Pharmacol.* **12**, 642820 (2021).
45. Y. Wang, F. Zhang, Y. Zhang, Q. Shan, W. Liu, F. Zhang, F. Zhang, S. Yi, Betacellulin regulates peripheral nerve regeneration by affecting Schwann cell migration and axon elongation. *Mol. Med.* **27**, 27 (2021).
46. A. Wangzhou, C. Paige, S. V. Neerukonda, D. K. Naik, M. Kume, E. T. David, G. Dussor, P. R. Ray, T. J. Price, A ligand-receptor interactome platform for discovery of pain mechanisms and therapeutic targets. *Sci. Signal.* **14**, abe1648 (2021).
47. Z. Kee, X. Kodji, S. D. Brain, The role of calcitonin gene related peptide (CGRP) in neurogenic vasodilation and its cardioprotective effects. *Front. Physiol.* **9**, 1249 (2018).
48. D. Aletaha, T. Neogi, A. J. Silman, J. Funovits, D. T. Felson, C. O. Bingham III, N. S. Birnbaum, G. R. Burmester, V. P. Bykerk, M. D. Cohen, B. Combe, K. H. Costenbader, M. Dougados, P. Emery, G. Ferraccioli, J. M. W. Hazes, K. Hobbs, T. W. J. Huizinga, A. Kavanaugh, J. Kay, T. K. Kvien, T. Laing, P. Mease, H. A. Ménard, L. W. Moreland, R. L. Naden, T. Pincus, J. S. Smolen, E. Stanislawska-Biernat, D. Symmons, P. P. Tak, K. S. Upchurch, J. Vencovský, F. Wolfe, G. Hawker, 2010 rheumatoid arthritis classification criteria: An American College of Rheumatology/European League Against Rheumatism collaborative initiative. *Arthritis Rheum.* **62**, 2569–2581 (2010).

49. F. C. Arnett, S. M. Edworthy, D. A. Bloch, D. J. McShane, J. F. Fries, N. S. Cooper, L. A. Healey, S. R. Kaplan, M. H. Liang, H. S. Luthra, The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. *Arthritis Rheum.* **31**, 315–324 (1988).
50. E. M. Roos, H. P. Roos, L. S. Lohmander, C. Ekdahl, B. D. Beynnon, Knee Injury and Osteoarthritis Outcome Score (KOOS)—Development of a self-administered outcome measure. *J. Orthop. Sports Phys. Ther.* **28**, 88–96 (1998).
51. A. K. Nilsdotter, L. S. Lohmander, M. Klässbo, E. M. Roos, Hip disability and osteoarthritis outcome score (HOOS) – Validity and responsiveness in total hip replacement. *BMC Musculoskelet. Disord.* **4**, 1–8 (2003).
52. A. Dobin, C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, T. R. Gingeras, STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
53. J. T. Leek, J. D. Storey, Capturing heterogeneity in gene expression studies by surrogate variable analysis. *PLOS Genet.* **3**, 1724–1735 (2007).
54. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
55. Y. Saeys, I. Inza, P. Larrañaga, A review of feature selection techniques in bioinformatics. *Bioinformatics* **23**, 2507–2517 (2007).
56. S. Liu, C. Xu, Y. Zhang, J. Liu, B. Yu, X. Liu, M. Dehmer, Feature selection of gene expression data for cancer classification using double RBF-kernels. *BMC Bioinformatics* **19**, 396 (2018).
57. F. Uzma, A. Al-Obeidat, B. Tubaishat, Z. Shah, Gene encoder: A feature selection technique through unsupervised deep learning-based clustering for large gene expression data. *Neural Comput. Appl.* **34**, 8309–8331 (2022).
58. P. Dhal, C. Azad, A comprehensive survey on feature selection in the various fields of machine learning. *Appl. Intell.* **52**, 4543–4581 (2022).
59. M. S. D. Koller, Toward optimal feature selection, in *ICML'96: Proceedings of the Thirteenth International Conference on International Conference on Machine Learning*, L. Saitta, Ed. (Morgan Kaufmann Publishers Inc., 1996), pp. 284–292.
60. H. Wang, T. M. Khoshgoftaar, K. Gao, A comparative study of filter-based feature ranking techniques, in *2010 IEEE International Conference on Information Reuse & Integration* (IEEE, 2010), pp. 43–48.
61. J. C. Ang, A. Mirzal, H. Haron, H. N. A. Hamed, Supervised, unsupervised, and semi-supervised feature selection: A review on gene selection. *IEEE/ACM Trans. Comput. Biol. Bioinformatics* **13**, 971–989 (2013).
62. S. Solorio-Fernández, J. A. Carrasco-Ochoa, J. F. Martínez-Trinidad, A review of unsupervised feature selection methods. *Artif. Intell. Rev.* **53**, 907–948 (2019).
63. N. El Aboudi, L. Benhlima, Review on wrapper feature selection approaches, in *2016 International Conference on Engineering & MIS (ICEMIS)* (IEEE, 2016), pp. 1–5.
64. M. Gönen, E. Alpaydin, Multiple kernel learning algorithms. *J. Mach. Learn. Res.* **12**, 2211–2268 (2011).
65. L. van der Maaten, G. Hinton, Visualizing data using t-SNE. *J. Mach. Learn. Res.* **9**, 2579–2605 (2008).
66. B. Wang, J. Zhu, E. Pierson, D. Ramazzotti, S. Batzoglou, Visualization and analysis of single-cell RNA-seq data by kernel-based similarity learning. *Nat. Methods* **14**, 414–416 (2017).
67. G. E. Hinton, S. Roweis, Stochastic neighbor embedding, in *Proceedings of the 15th International Conference on Neural Information Processing Systems* (MIT Press, 2002), pp. 857–864.
68. J. Reimand, M. Kull, H. Peterson, J. Hansen, J. Vilo, g:Profiler—A web-based toolset for functional profiling of gene lists from large-scale experiments. *Nucleic Acids Res.* **35**, W193–W200 (2007).
69. G. Korotkevich, V. Sukhov, N. Budin, B. Shpak, M. N. Artyomov, A. Sergushichev, Fast gene set enrichment analysis. bioRxiv 060012 [Preprint] (2021); https://doi.org/10.1101/060012.
70. U. Raudvere, L. Kolberg, I. Kuzmin, T. Arak, g:Profiler: A web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res.* **47**, W191–W198 (2019).
71. M. Kanehisa, S. Goto, Y. Sato, M. Furumichi, M. Tanabe, KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res.* **40**, D109–D114 (2012).
72. A. Fabregat, S. Jupe, L. Matthews, K. Sidiropoulos, M. Gillespie, P. Garapati, R. Haw, B. Jassal, F. Korninger, B. May, M. Milacic, C. D. Roca, K. Rothfels, C. Sevilla, V. Shamovsky, S. Shorser, T. Varusai, G. Viteri, J. Weiser, G. Wu, L. Stein, H. Hermjakob, P. D'Eustachio, The reactome pathway knowledgebase. *Nucleic Acids Res.* **46**, D649–D655 (2018).
73. T. Kelder, M. P. van Iersel, K. Hanspers, M. Kutmon, B. R. Conklin, C. T. Evelo, A. R. Pico, WikiPathways: Building research communities on biological pathways. *Nucleic Acids Res.* **40**, D1301–D1307 (2012).
74. M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin, G. Sherlock, Gene Ontology: Tool for the unification of biology. *Nat. Genet.* **25**, 25–29 (2000).
75. J. A. Ramilowski, T. Goldberg, J. Harshbarger, E. Kloppmann, M. Lizio, V. P. Satagopam, M. Itoh, H. Kawaji, P. Carninci, B. Rost, A. R. R. Forrest, A draft network of ligand–receptor-mediated multicellular signalling in human. *Nat. Commun.* **6**, 1–12 (2015).
76. J. X. Binder, S. Pletscher-Frankild, K. Tsafou, C. Stolte, S. I. O'Donoghue, R. Schneider, L. J. Jensen, COMPARTMENTS: Unification and visualization of protein subcellular localization evidence. *Database* **2014**, bau012 (2014).
77. B. Braschi, P. Denny, K. Gray, T. Jones, R. Seal, S. Tweedie, B. Yates, E. Bruford, Genenames.org: The HGNC and VGNC resources in 2019. *Nucleic Acids Res.* **47**, D786–D792 (2019).
78. S. Carbon, A. Ireland, C. J. Mungall, S. Shu, B. Marshall, S. Lewis, AmiGO Hub, Web Presence Working Group, AmiGO: Online access to ontology and annotation data. *Bioinformatics* **25**, 288–289 (2009).
79. D. Szklarczyk, A. L. Gable, D. Lyon, A. Junge, S. Wyder, J. Huerta-Cepas, M. Simonovic, N. T. Doncheva, J. H. Morris, P. Bork, L. J. Jensen, C. von Mering, STRING v11: Protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **47**, D607–D613 (2018).
80. B. Yates, B. Braschi, K. A. Gray, R. L. Seal, S. Tweedie, E. A. Bruford, Genenames.org: The HGNC and VGNC resources in 2017. *Nucleic Acids Res.* **45**, D619–D625 (2017).
81. Y. Hao, S. Hao, E. Andersen-Nissen, W. M. Mauck III, S. Zheng, A. Butler, M. J. Lee, A. J. Wilk, C. Darby, M. Zager, P. Hoffman, M. Stoeckius, E. Papalexi, E. P. Mimitou, J. Jain, A. Srivastava, T. Stuart, L. M. Fleming, B. Yeung, A. J. Rogers, J. M. McElrath, C. A. Blish, R. Gottardo, P. Smibert, R. Satija, Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587.e29 (2021).
82. P. D. Thomas, D. Ebert, A. Muruganujan, T. Mushayahama, L.-P. Albou, H. Mi, PANTHER: Making genome-scale phylogenetics accessible to all. *Protein Sci.* **31**, 8–22 (2022).
83. A. Zeisel, H. Hochgerner, P. Lönnerberg, A. Johnsson, F. Memic, J. van der Zwan, M. Häring, E. Braun, L. E. Borm, G. La Manno, S. Codeluppi, A. Furlan, K. Lee, N. Skene, K. D. Harris, J. Hjerling-Leffler, E. Arenas, P. Ernfors, U. Marklund, S. Linnarsson, Molecular architecture of the mouse nervous system. *Cell* **174**, 021 (2018).

Acknowledgments: We thank the participants who provided synovial tissue and blood samples for this study. We thank A. North and V. Sharma from the Rockefeller University's Bio-Imaging Resource Center, RRID:SCR_017791, for help with imaging and image analysis. **Funding:** This project was funded by the National Science Foundation 1750326 (to Z.B. and F.W.); the Arthritis Foundation (to S.G.); the NIH R01AR078268 (to D.E.O.), UC2AR081025 (to D.E.O. and S.G.), R01AR077019 (to R.E.M.), R01AR064251 (to A.-M.M.), R01AR060364 (to A.-M.M.), P30AR079206 (to A.-M.M.), and U19-NS130608 (to T.J.P.); Rockefeller University UL1TR001866 (to D.E.O.); Arthritis National Research Foundation Award (to F.Z.), UK Medical Research Council G0800648 (to C. Pitzalis), Versus Arthritis 20022 (to C. Pitzalis); FOREUM 232719 (to S.S.); EFIC Grünenthal Grant (to S.S.); Versus Arthritis 21734 (to S.S.); and the Accelerating Medicines Partnership Program: Rheumatoid Arthritis and Systemic Lupus Erythematosus (AMP RA/SLE) Network. The AMP Program is a public-private partnership that includes AbbVie Inc., the Arthritis Foundation, Bristol-Myers Squibb Company, the Foundation for the National Institutes of Health, GlaxoSmithKline, Janssen Research and Development LLC, the Lupus Foundation of America, the Lupus Research Alliance, Merck Sharp & Dohme Corp., the National Institute of Allergy and Infectious Diseases, the National Institute of Arthritis and Musculoskeletal and Skin Diseases, Pfizer Inc., the Rheumatology Research Foundation, Sanofi, and Takeda Pharmaceuticals International Inc. Funding for AMP RA/SLE work was provided through grants from the National Institutes of Health (UH2-AR067676, UH2-AR067677, UH2-AR067679, UH2-AR067681, UH2-AR067685, UH2-AR067688, UH2-AR067689, UH2-AR067690, UH2-AR067691, UH2-AR067694, and UM2-AR067678). **Author contributions:** Z.B., N.B., M.A., A.-M.M., T.J.P., R.E.M., S.G., R.B.D., M.O.F., A.-M.M., R.E.M., F.W., and D.E.O. conceptualized the study. E.S., E.D., S.G., and M.H.S. collected the patient clinical data and samples. Z.B. designed and implemented the algorithms and frameworks. Z.B., C.R.H., M.J.L., C. Pyrgaki, and S.S. analyzed the validation dataset. Z.B. analyzed the AMP data. T.J.P., K.M., and J.B.L. performed the receptor ligand interaction analysis. M.A., M.H.S., and N.E.B. tested the effects of synovial fibroblast products on neurons in vitro. E.A.M., N.E.B., S.P., C. Pyrgaki, R.E.M., and A.-M.M. performed and analyzed the tissue cleared immunofluorescent imaging data. Z.B., N.B., M.A., C.R.H., E.A.M., N.E.B., S.P., E.S., E.D., M.H.S., M.O.F., C.S.J., H.Z., M.J.L., S.S., C. Pitzalis, A.-M.M., J.B.L., K.M., T.J.P., R.E.M., F.Z., S.G., F.W., and D.E.O. analyzed the data. Z.B., N.B., M.A., E.A.M., C.R.H., N.E.B., S.P., E.S., E.D., E.M.G., M.H.S., M.O.F., C.S.J., H.Z., C. Pyrgaki, M.J.L., S.S., C. Pitzalis, J.B.L., K.M., T.J.P., A.-M.M., R.E.M., F.Z., S.G., R.B.D., F.W., and D.E.O. wrote and edited the manuscript. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data associated with this study are present in the paper or the Supplementary Materials. The bulk synovial RNA-seq dataset is available at ImmPort (SDY1299). The PEAC dataset is available at https://peac.hpc.qmul.ac.uk. The AMP dataset is available at https://immunogenomics.io/ampsle1/ or ImmPort (SDY997). The raw sequence data are deposited in the sequence read archive under accession SRP135960 in the National Center for Biotechnology Information (NCBI) library. Our code is available at DOI: 10.5281/zenodo.10783506 and target URL https://zenodo.org/doi/10.5281/zenodo.10783505. **AMP RA/SLE Network:** In addition to AMP RA/SLE Network members who are authors, the following AMP RA/SLE Network members are collaborators who have contributed to investigation and data curation of this manuscript:

Jennifer Albrecht[12], Jennifer H. Anolik[12], William Apruzzese[4], Brendan F. Boyce[12], David L. Boyle[13], Michael B. Brenner[4], S. Louis Bridges Jr.[3], Christopher D. Buckley[14], Jane H. Buckner[15], Vivian P. Bykerk[1,3], James Dolan[4], Laura T. Donlin[1,3], Andrew Filer[14], Gary S. Firestein[13], Chamith Y. Fonseka[4], Peter K. Gregersen[16], Joel M. Guthridge[17], Maria Gutierrez-Arcelus[4], V. Michael Holers[9], Laura B. Hughes[18], Lionel B. Ivashkiv[1,3], Eddie A. James[15], Judith A. James[17], A.Helena Jonsson[9,4], Stephen Kelly[19], James A. Lederer[4], Yvonne C. Lee[20], Arthur M. Mandelin II[20], Mandy J. McGeachy[21], Joseph R. Mears[4], Nida Meednu[12], Larry Moreland[9,21], Harris Perlman[20], Javier Rangel-Moreno[12], Deepak A. Rao[4], Soumya Raychaudhuri[4], Christopher Ritchlin[12], William H. Robinson[22], Mina Rohani-Pichavant[22], Jennifer Seifert[9], Kamil Slowikowski[4], Darren Tabechian[12], Paul J. Utz[22], Gerald F. M. Watts[4], Kevin Wei[4]

[12]University of Rochester Medical Center, Rochester, NY 14642, USA. [13]University of California, San Diego, La Jolla, CA 92093, USA. [14]University Hospitals Birmingham NHS Foundation Trust and University of Birmingham, Birmingham B15 2GW, UK. [15]Benaroya Research Institute at Virginia Mason, Seattle, WA 98101-2795, USA. [16]Feinstein Institute for Medical Research, Northwell Health, Manhasset, New York, NY 11030, USA. [17]Oklahoma Medical Research Foundation, Oklahoma City, OK 73104, USA. [18]University of Alabama at Birmingham, Birmingham, AL 35294, USA. [19]Barts Health NHS Trust, London E1 2ES, UK. [20]Northwestern University Feinberg School of Medicine, Chicago, IL 60611, USA. [21]University of Pittsburgh School of Medicine, Pittsburgh, PA 15261, USA. [22]Stanford University School of Medicine, Palo Alto, CA, 94305.

Submitted 17 August 2023
Resubmitted 2 January 2024
Accepted 21 March 2024
Published 10 April 2024
10.1126/scitranslmed.adk3506