Genome-wide screens identify SEL1L as an intracellular rheostat controlling collagen turnover

Michael J. Podolsky[1,*], Benjamin Kheyfets[1], Afaq H. Beigh[1], Christopher D. Yang[2], Carlos Lizama Valenzuela[2], Ritwik Datta[2], Paul J. Wolters[3], Michael T. McManus[4], Ling Qi[5,6], Kamran Atabai[2,3,7,*]

*Co-corresponding author

1- Department of Medicine, Weill Cornell Medical College, New York, NY, USA.
2- Cardiovascular Research Institute, University of California, San Francisco, CA, USA.
3- Department of Medicine, University of California, San Francisco, CA, USA.
4- Department of Microbiology and Immunology and UCSF Diabetes Center, University of California, San Francisco, CA, USA.
5- Department of Internal Medicine, Division of Metabolism, Endocrinology and Diabetes, University of Michigan Medical School, Ann Arbor, MI, USA.
6- Department of Molecular and Integrative Physiology, University of Michigan Medical School, Ann Arbor, MI, USA.
7- Lung Biology Center, University of California, San Francisco, CA, USA.

The authors have declared that no conflict of interest exists.

Abstract

Accumulating evidence has implicated impaired extracellular matrix (ECM) clearance as a key factor in fibrotic disease. Despite decades of research elucidating the effectors of ECM

clearance, relatively little is understood regarding the upstream regulation of this process. Collagen is the most abundant constituent of normal and fibrotic ECM in mammalian tissues. Its catabolism occurs through extracellular proteolysis and cell-mediated uptake of collagen fragments for intracellular degradation. Given the paucity of information regarding the regulation of this latter process, we executed unbiased genome-wide screens to understand the molecular underpinnings of cell-mediated collagen clearance. Using this approach, we discovered a previously unappreciated mechanism through which collagen biosynthesis is sensed by cells internally and directly regulates clearance of extracellular collagen. The sensing mechanism is dependent on endoplasmic reticulum-resident protein SEL1L and occurs via a noncanonical function of SEL1L. This pathway functions as a homeostatic negative feedback loop that limits collagen accumulation in tissues. In human fibrotic lung disease, the induction of this collagen clearance pathway by collagen synthesis is impaired, thereby contributing to the pathological accumulation of collagen in lung tissue. Thus cell-autonomous, rheostatic collagen clearance is a previously unidentified pathway of tissue homeostasis.

Main Text

Fibrosis can be seen as the excess accumulation of extracellular matrix that interferes with tissue and organ function. Collagen is the most abundant constituent of fibrotic extracellular matrix, yet we still lack a complete understanding of the regulatory mechanisms governing collagen homeostasis. Whereas decades of research have elucidated pathways of collagen deposition in normal tissue and in fibrosis, relatively less is understood about collagen clearance. Clearance of extracellular collagen from tissue occurs via two pathways that can act in parallel and sequentially[1]: an extracellular proteolytic process and a process of intracellular uptake and degradation. This latter process is vitally important in controlling normal tissue

53 homeostasis and mitigating fibrosis as demonstrated in several recent publications that show
54 the effects of interfering with this process in animal models of fibrosis[2,3]. Although an impaired
55 collagen-degradative environment is a feature of fibrotic diseases such as Idiopathic Pulmonary
56 Fibrosis (IPF), Scleroderma, and cirrhosis[4–8], very little is known about the role of cell-based
57 collagen degradation in human disease.

58 A key mediator of cell-based collagen clearance is MRC2, a canonical collagen
59 endocytic receptor that binds to and internalizes collagen via a Fibronectin-2 domain, with
60 subsequent lysosomal degradation[9,10]. Earlier studies of this receptor showed that it mitigates
61 lung fibrosis[3] as well as liver and kidney models of fibrosis[11,12]. Our recent work added to this
62 literature by showing that MRC2-mediated collagen clearance is essential for resolution of
63 fibrosis (rather than development of fibrosis)[13]; deletion of MRC2 is sufficient to impair resolution
64 of fibrosis in the bleomycin mouse model of lung fibrosis.

65 We and others have been studying the process of cellular uptake and degradation of
66 collagen fragments because of the accumulating evidence that this is an important homeostatic
67 mechanism *in vivo* and controls fibrosis resolution[1–3,11–14]. Previous attempts at uncovering novel
68 mediators of this process have met with some success[15,16]. However, experimental approaches
69 have either been targeted based on known genes and pathways or have been limited to a
70 subset of potential genes because of limited technical ability to use genome wide discovery
71 approaches. The advent of high throughput CRISPR-based screens that are done in a pooled
72 format and leverage high throughput sequencing[17] has now permitted the unbiased evaluation
73 of all potential genetic regulatory pathways that could be contributing to this process. In this
74 work, we started with complementary CRISPRi and CRISPRa screens to address this
75 knowledge gap. Unexpectedly we have identified a rheostatic mechanism controlling collagen
76 clearance that was previously unidentified: collagen biosynthesis positively regulates cell-based
77 collagen clearance. In so doing, we found a novel role for SEL1L, an endoplasmic reticulum-

78 resident protein, which behaves as an internal cellular sensor of collagen biosynthesis and
79 controls cell-based collagen clearance.
80
81
82 Results
83
84 We performed genome wide unbiased screens using CRISPR interference (CRISPRi)
85 and CRISPR activation (CRISPRa) to determine genes that regulate cellular uptake of
86 exogenous collagen (Fig. 1). In independent pooled screens, we used CRISPRi to silence every
87 gene in the genome, or CRISPRa to activate every gene in the genome along with use of
88 10,000 non-targeting guides. We then performed a phenotypic screen for cellular uptake of
89 fluorescent collagen fragments. The screens were carried out independently with entirely
90 separate guide libraries for CRISPRi and CRISPRa. Coverage was at least 350x throughout the
91 screening process for each screen. The screens were performed in U937 cells, a human
92 monocytic leukemia cell line that readily takes up collagen[18] and grows in suspension, which
93 facilitates the use of a large-scale screen platform requiring culturing of hundreds of millions of
94 cells.
95 We determined multiple genetic regulators of collagen uptake (Fig. 1B-C; Supp. Table
96 1). Some of these were previously known to have a role in this process, most importantly
97 *MRC2*, a canonical collagen endocytic receptor. Most of the other candidates were not
98 previously known to play a role in collagen turnover. Importantly, we found a negative
99 correlation when comparing phenotype scores for individual genes that were hits in both the
100 CRISPRi and CRISPRa screens (Fig. 1D). This suggests that at least some of these genes we
101 identified are both necessary and sufficient for cellular collagen clearance and that quantitative
102 control of collagen clearance by these genes might be tunable based on their expression levels.
103 To validate our findings, we re-screened the high and low uptake bins from the original CRISPR

104 screens individually and compared hits in these re-screens to the corresponding original
105 CRISPRi or CRISPRa screen (Supp. Fig. 1A-D). We found a marked correlation between the
106 original and repeat screens (Supp. Fig. 1C-D). To validate the performance of our pooled guide
107 library, we used arrayed delivery of CRISPRi reagents for 3 sgRNAs for multiple genes
108 identified as hits in either positive or negative directions in the original CRISPRi screens,
109 followed by collagen uptake assays. Phenotypic effects of arrayed delivery of these CRISPRi
110 sgRNAs were 77% concordant with original screen data (Supp. Fig. 1E).

111 We next performed bioinformatic analyses of hits from our CRISPR screens. We noticed
112 that multiple gene ontology (GO) groups related to collagen metabolism were top scoring GO
113 terms in our analysis using GSEA (Fig. 2A). This was true in all four screens, i.e. the original
114 CRISRPi and CRISPRa screens as well as the re-screens (Supp. Fig. 2). Intriguingly, this did
115 not only include collagen catabolism GO terms, but also collagen biosynthesis or collagen
116 production-related GO terms (Fig. 2A). This informatic analysis suggested that collagen
117 biosynthetic activity positively regulates cellular uptake of extracellular collagen fragments. We
118 compared the enrichment of these GO terms in our collagen uptake screens with data from
119 previously published CRISPR screens of phagocytosis of other cargo[19] (Fig. 2B). Only our
120 screens for collagen uptake exhibited uniform enrichment of collagen biosynthetic GO terms
121 (i.e. positive regulation of collagen uptake by collagen biosynthesis), compared with uptake of
122 other cargo (Fig. 2B). These data indicate that there is a biological basis for positive regulation
123 collagen uptake by collagen biosynthesis genes, and this is specific to uptake of collagen.

124 To validate whether collagen biosynthesis positively regulates collagen uptake, we
125 decided to evaluate individual collagen biosynthesis genes. A heat map of gene-level phenotype
126 scores from our original screens among top-enriched genes from the Collagen Biosynthesis GO
127 group is shown in Figure 3A. We also note that *TRAM2*, a gene previously described to be
128 critically important collagen biosynthesis[20] was a top hit in the CRISPRi screen as well (Fig. 1).
129 We used an orthogonal technique (shRNA) to silence TRAM2 in U937 cells which resulted in

130 decreased collagen uptake (Fig. 3B-C) and decreased MRC2 expression at the protein and
131 mRNA levels (Fig. 3D-E). To verify whether TRAM2-silencing indeed affects collagen synthesis,
132 we used MRC5 fibroblasts which express high levels of collagen. Silencing of TRAM2 in
133 fibroblasts led to decreased endogenous levels of collagen (Fig. 3F-G) and decreased collagen
134 uptake in fibroblasts as well (Fig. 3H). We next silenced individual collagen genes that were
135 seen in the original screen including type I collagens and type V collagens and determined that
136 silencing of these genes led to decreased collagen uptake and a decrease in *MRC2* message in
137 U937 cells. By contrast, stimulating collagen expression by pharmacologic means with TGFβ1
138 increased collagen uptake in U937 cells (Fig. 3L). Although it is controversial whether
139 leukocytes or myeloid cells produce significant quantities of collagen[21–23], they do generate
140 collagen mRNA and recently were found to contribute to collagen content *in vivo* under certain
141 conditions[24]; a Western blot verified that U937 cells do indeed make collagen, albeit in very low
142 quantities (Fig. 3L; note 60 μg of protein needed to be loaded to visualize collagen from U937
143 cells, compared with 20 μg loaded in most other Western blots in this manuscript; additionally, a
144 more sensitive total type I collagen antibody was used in this subpanel rather than a procollagen
145 antibody that is less sensitive but is specific for newly synthesized collagen, which we used in
146 the other subpanels of this figure). To corroborate these findings, we additionally stimulated
147 collagen synthesis in MRC5 fibroblasts and measured collagen uptake and *MRC2* expression.
148 Stimulation with TGFβ1 led to increased collagen uptake and increased *MRC2* expression (Fig.
149 3M-N). Using orthogonal pharmacologic stimulators of collagen synthesis, ascorbic acid and
150 LPA1 respectively, led to increasing *MRC2* message as well (Fig. 3O-P). These data together
151 suggest a conserved pathway in which collagen biosynthesis directly regulates *MRC2*
152 expression and collagen uptake.
153 Because the screens were based on flow cytometric sorting of individual cells and each
154 cell had a single individual genetic perturbation, phenotypes determined by our screens are

155 expected to be fundamentally cell-autonomous. Therefore we presumed that collagen synthesis
156 in cells must be sensed by some internal cell-autonomous mechanism. In examining hits in the
157 screen to determine whether there could be a cell-autonomous sensor of collagen synthesis, we
158 were intrigued to see that *SEL1L* was a top hit in our screens (Fig. 1). The gene product of
159 *SEL1L* is a single-pass, endoplasmic reticulum-resident transmembrane protein with its luminal
160 face containing a Fibronectin-2 (FN2) domain (Fig. 4A). FN2 domains are generally known to
161 bind collagen[25–27]. However the function of the FN2 domain in SEL1L has not previously been
162 characterized.

163 To validate the effect of *SEL1L* on collagen uptake, we used an inducible knockout
164 mouse embryonic fibroblast cell line[28]. We determined that when *Sel1L* is deleted, there is
165 reduced collagen uptake (Fig. 4B), reduced MRC2 protein expression (Fig. 4C), reduced *Mrc2*
166 mRNA (Fig. 4D), and reduced MRC2 cell surface and total protein expression (Fig. 4E-F). By
167 contrast overexpressing SEL1L in fibroblasts leads to increased *MRC2* message and protein
168 (Fig. 4G-H). Importantly although SEL1L has chiefly been described as a member of the
169 endoplasmic reticulum-associated protein degradation (ERAD) quality control pathway[28–30],
170 other members of the ERAD pathway were not hits in our CRISPR screens of collagen uptake
171 (Fig. 1; Supp. Table 1). This included HRD1, which acts with SEL1L. Furthermore, to validate
172 SEL1L-mediated regulation of collagen uptake is not ERAD-dependent, we used cells null for
173 SEL1L or HRD1, and determined that only SEL1L positively regulates collagen uptake, MRC2
174 expression (Fig. 5A-B), and *MRC2* message production (Fig. 5C). Furthermore, MRC2 can be
175 forcibly overexpressed in wildtype (WT), SEL1L-knockout and HRD1-knockout cells equally with
176 resulting MRC2 that is functional and can take up exogenous fluorescent collagen fragments
177 (Fig. 5D-F). This indicates that the ERAD pathway is not necessary for MRC2 protein
178 expression or maturation and argues against ERAD itself being instrumental in regulating MRC2
179 expression. Furthermore, our data indicate that *MRC2* is being regulated by SEL1L at the

180 message level based on the data in Figures 4 and 5, rather than protein level as may be
181 expected if the process were ERAD-dependent.
182 To determine whether SEL1L could be a collagen sensor, we wanted to investigate
183 whether it binds collagen and whether its FN2 domain is important for its function of regulating
184 *MRC2* levels. In fibroblasts, we determined that SEL1L and collagen co-localize, and this
185 colocalization increases when collagen synthesis is induced by TGFβ1 (Fig. 6A). Furthermore,
186 there is a biochemical association between SEL1L and type I collagen, as determined by co-
187 immunoprecipitation followed by Western blotting (Fig. 6B). We next made expression
188 constructs of WT SEL1L vs. SEL1L with an in-frame deletion of its FN2 domain (AA118-166;
189 referred to as ΔFN2) to examine the function of this domain. In fact, the SEL1LΔFN2 mutant
190 exhibited markedly diminished binding to procollagen compared to WT SEL1L (Fig. 6C-D).
191 Importantly in both fibroblasts and HEK293T cells, WT SEL1L drives MRC2 expression but
192 SEL1LΔFN2 does not, even though in these experiments these expression constructs are
193 expressed at similar levels (Fig. 6E-G). Taken together these data strongly suggest that SEL1L
194 is an internal sensor of collagen biosynthesis, and SEL1L is necessary for the effect of collagen
195 biosynthesis on *MRC2* message levels and hence the downstream phenotype of collagen
196 uptake. Consistent with our data in Figure 5 showing this phenomenon is ERAD-independent, a
197 comparison of protein domain structures of SEL1L across diverse species demonstrates that
198 although SEL1L is evolutionarily ancient and conserved down to yeast[31], the FN2 domain is only
199 present in vertebrates and higher organisms, showing the collagen-sensing function arose later
200 and is likely to be independent of its original ERAD function (Fig. 6H).
201 To determine whether this pathway is relevant in human physiology and disease, we re-
202 analyzed single cell data from multiple published experiments including normal and fibrotic lung.
203 We found that in cells that express both *MRC2* and type I collagen, there is a positive
204 correlation between expression levels of these two mRNA species (Fig. 7A-D). However, this

205 positive correlation is diminished IPF lung cells (Fig. 7D). In fact, data from multiple independent
206 diverse datasets examining gene expression in different compartments (whole lung,
207 bronchoalveolar lavage, peripheral blood mononuclear cells) from control versus IPF patients
208 demonstrate that in all cases there is a positive correlation between collagen and *MRC2*
209 expression, but the relationship is impaired in IPF (Fig. 7E-G). These findings corroborate our *in*
210 *vitro* data showing collagen biosynthesis is positively associated with *MRC2* and collagen
211 uptake by cells. This diminished recruitment of *MRC2* could be a potential mechanism by which
212 fibrotic collagenous extracellular matrix fails to be cleared in diseases of persistent and
213 progressive fibrosis like IPF. By contrast in a spontaneously resolving model of fibrosis (the
214 bleomycin lung fibrosis model in C57Bl\6 mice) there is a positive correlation between collagen
215 and *Mrc2* expression in mouse lung and the relationship is not impaired under conditions of
216 fibrosis (Fig. 7H). Given that we have previously shown that MRC2 is necessary for resolution of
217 fibrosis[13], these data indicate that the impairment of biosynthesis-induced collagen clearance
218 via MRC2 could be an important pathway driving non-resolving fibrosis and IPF.
219 Finally, we examined SEL1L levels in IPF lung versus control lung. We found that in
220 some, but not all, IPF lung specimens there is markedly diminished full-length SEL1L (Fig. 8A-
221 C). Interestingly, this appears to be primarily at the protein level and not at the mRNA level (Fig.
222 8D). Future work will investigate how this process is taking place. Whether this is the only
223 mechanism or one of the mechanisms by which collagen biosynthesis-induced collagen
224 clearance is impaired in IPF is also an area of study in our laboratory.
225
226
227 Discussion
228
229 In this work we have shown that collagen synthesis directly upregulates collagen
230 clearance by MRC2. Our screens identified uniquely important biological pathways controlling

231 collagen clearance that do not regulate phagocytic uptake of other cargo. This led us to identify
232 SEL1L as a critical regulator of collagen synthesis-induced collagen turnover. SEL1L is an
233 internal sensor of collagen biosynthesis and its collagen sensing function is necessary for the
234 homeostatic effect of collagen biosynthesis on upregulation of MRC2. SEL1L is known to be a
235 key member of the ERAD pathway, but our data indicate that the collagen-sensing function of
236 SEL1L is independent of its ERAD function, defining a novel non-canonical function for this
237 protein. Consistent with this concept is our observation that although SEL1L and the ERAD
238 pathways are both evolutionarily ancient with conservation to yeast[31], the collagen-sensing FN2
239 domain of SEL1L arose more recently with vertebrates after the evolutionary origin of collagen
240 proteins, which appeared concurrently with metazoa[32].

241 From a teleological perspective, priming the same cells that contribute to scar formation
242 to upregulate machinery integral for repair and clearance of that scar would be an efficient
243 means of reaching a goal of tissue homeostasis and regeneration of normal tissue. It is striking
244 that in IPF, a disease of persistent non-resolving fibrosis, the relationship between collagen
245 synthesis and clearance is uncoupled, in contrast to the preserved relationship of these two
246 processes in the spontaneously resolving model of bleomycin-induced lung fibrosis in C57Bl\6
247 mice[13,33,34]. These findings suggest this pathway could be ripe for therapeutic intervention in
248 IPF. Until now, it has not been apparent that this rheostatic collagen turnover pathway was
249 inadequately engaged because previous studies have shown a modest increase in *MRC2*
250 expression in IPF lung compared with normal lung[35–39]. However, this belies a dramatically
251 inappropriate recruitment of the homeostatic turnover mechanism necessary to resolve fibrosis
252 which would be expected commensurate with high levels of collagen synthesis in IPF.

253 The mechanism by which SEL1L-mediated collagen sensing transcriptionally regulates
254 *MRC2* is not completely determined by this work. The luciferase activity data we have produced
255 suggest that *MRC2* transcript production may be a downstream effect of SEL1L expression
256 levels. The precise transcription factors or regulators that are responsible for this are the subject

257 of ongoing study in our laboratory. We also do not yet understand how SEL1L is downregulated
258 in IPF. Our data suggest this is one potential mechanism by which there is incomplete
259 recruitment of the homeostatic collagen clearance mechanisms in IPF. How SEL1L itself is
260 downregulated in IPF issue is also the subject of ongoing study in our laboratory. Nevertheless,
261 we demonstrate here that unbiased genome-wide approaches based on phenotypic screening
262 can yield new insights into cell and tissue-level biology.
263
264
265 Methods
266
267 **CRISPR Screens**
268 Single cell clones of U937 cells expressing dCas9-KRAB-BFP (CRISPRi) or dCas9-VP64-BFP
269 (CRISPRa) were made as previously described[40]. Subsequently, independent guide libraries for
270 CRISPRi and CRISPRa were used as previously described[41]. Briefly, lentivirus particles
271 containing these guides were transduced into U937 clones above at low MOI (0.3), followed by
272 subsequent selection with puromycin (1 µg/ml). Puromycin-selected cells were amplified and the
273 underwent fluorescent collagen uptake assay (as below) in bulk (i.e. pooled format), with sorting
274 into high and low bins (top or bottom 10% of collagen uptake based on mean fluorescence
275 intensity), simultaneously on 4x FACSAria II devices, then re-cultured and harvested for
276 analysis or for repeat screening. Isolation of cell DNA, followed by amplification of the guide-
277 region of individual cells and high-throughput amplicon sequencing on an Illumina platform were
278 all done as previously described[41,42].
279
280 **Bioinformatics, Including Analysis of Screen Data**
281 Guide counts were extracted using Python v3 scripts previously published[17]. Analysis of screen
282 data was done via MAGeCK[43] using the lfc-mean method. Bioinformatic analysis was performed

283 using GSEA[44,45] on phenotype scores (defined as Log-fold change multiplied by the $-\log_{10}$(p-
284 value) as determined by the robust rank-aggregation method in MAGeCK). For comparison
285 across knockdown screens for different cargo, data from Haney MS et al. were used[19] and then
286 re-analyzed via GSEA along with our own collagen screens. Visualization of screen data was
287 done with ggplot2 package in R (version 4.1.1). Visualization of domain architecture of SEL1L
288 protein across species was performed using the OMA Orthology Browser[46]. For re-analysis of
289 single cell RNAseq data from multiple experiments (GSE136831, GSE135893, GSE121611,
290 GSE128033, GSE132771) were integrated and re-normalized via the Seurat R package version
291 4; cells with non-zero expression of *COL1A1* and *MRC2* were used for analysis in Fig. 7D.
292 Microarray data was re-analyzed as indicated using GEO2R.
293
294 **Quantitative real-time PCR**
295 Total RNA from tissue or sorted cells was isolated via Trizol (Invitrogen) or RNeasy kit (Qiagen).
296 Total RNA from each sample was reverse transcribed with a first-strand cDNA synthesis kit
297 (Quantabio or Thermo Fisher) according to the manufacturer's instructions. Quantitative real-
298 time PCR reactions were performed with different sets of primers and Sensifast SYBR (Bioline,
299 Taunton, MA) on a CFX96 or CFX384 Real-Time PCR Detection System (Bio-Rad) or Applied
300 Biosystems QuantStudio 5 using 20 ng cDNA/reaction. Each quantitative real-time PCR
301 reaction was performed at least twice and representative results are shown relative to a control
302 by the standard $2^{-\Delta\Delta Ct}$ method, where Ct represents the number of cycles required to reach
303 threshold for the target gene subtracted from the number of cycles required to reach threshold
304 for a control housekeeping gene (*GAPDH* or *ACTB*). All data are shown relative to the first
305 control condition in the given figure panel. Primer sequences are in Supplemental Table 2.
306
307 **Immunoblotting**

308 Cells or tissues were homogenized with cold lysis buffer (RIPA buffer: 50 mM Tris-HCl pH 7.4,
309 150 mM NaCl, 0.1% SDS, 0.5% DOC, and 1% Nonidet P-40; coIP buffer: 20mM Tris-HCl pH 8,
310 137mM NaCl, 2mM EDTA, 1% Nonidet P-40) supplemented with protease and phosphatase
311 inhibitor cocktail (Thermo Fisher Scientific). Lysates (20 μg except where otherwise noted) were
312 clarified by centrifugation at 14,000g for 15 minutes at 4°C, electrophoresed under reducing
313 conditions on 7.5% or 4-20% gradient precast gels (Bio-Rad), and transferred to PVDF
314 membranes (Bio-Rad and GVS Filter Technology). Membranes were blocked for 30 minutes in
315 5% Bovine Serum Albumin and incubated with primary antibodies against MRC2 (1:1,200: anti-
316 mouse: AF4789, 1μg/mL: anti-human: AF5770; R&D Systems, Bio-Techne), SEL1L (1:2,000;
317 Novus Biologicals: NBP2-93746; ABCAM: ab78298), collagen I (1:2,000; Southern Biotech:
318 1310-01), human procollagen I alpha 1 (1:1,500; Novus Biologicals: AF6220), MYC (1:5,000;
319 Cell Signaling Technology, 9B11) GAPDH (1:4,000; Cell Signaling Technology, 14C10), or beta-
320 Actin (1:10,000; Cell Signaling Technology, 13E5) overnight at 4°C. Membranes were then
321 incubated with horseradish peroxidase–conjugated secondary antibodies (1:4000: Cell
322 Signaling Technology; 1:2,000: R&D Systems, Bio-Techne; Santa Cruz Biotechnology) for 1
323 hour at room temperature. Blots were developed using enhanced chemiluminescence (Cytiva)
324 by exposure onto film or ChemiDoc XRS+ (Bio-Rad). Where indicated densitometry was
325 performed using ImageJ (NIH), normalizing the protein of interest to a loading control (GAPDH
326 or beta-Actin), and comparisons were made to the baseline condition. For immunoblots of FN2-
327 deleted SEL1L, the Abcam antibody was used because the immunogen is distal to the FN2
328 domain, whereas the immunogen for the Novus antibody is the FN2 domain itself.
329
330 **Co-immunoprecipitation**
331 Proteins were isolated in non-denaturing coIP lysis buffer (see above). Proteins were incubated
332 with fast-flow protein G- or protein A- Dynabeads (Invitrogen) for 2 hours and centrifuged to
333 eliminate nonspecifically bound proteins. The concentrations of the precleared proteins were

334 estimated by Bradford's assay. Primary antibodies against Procollagen (Novus) or SEL1L
335 (Novus) or a non-specific IgG control antibody were incubated with protein G- or protein A-
336 Dynabeads for 4 hours. Then, 200-400 µg pre-cleared protein were incubated overnight with the
337 antibody-bound beads. After washing, attached proteins were then eluted from the beads in 1%
338 SDS buffer by heating at 60˚C and vortexing followed by centrifugation to collect the
339 supernatants as the eluates. Equal volumes of the eluates were used for immunoblotting.
340 Immunoprecipitation with protein isolation buffer without lysates was run as a negative control
341 for the beads. Proteins immunoprecipitated using non-specific IgG (NS IgG) were negative
342 control for specific antibodies. Immunoblotting with total pre-cleared protein lysates were used
343 as input samples.
344
345 **Flow cytometry and flow cytometric-based collagen uptake assay**
346 Flow cytometric analysis was done using a FACSVerse, FACSAriaII, or Accuri c6 (BD
347 Biosciences) and sorting was done using FACSAriaII devices (BD Biosciences) at the UCSF
348 Parnassus Flow Cytometry Core. For the collagen uptake assay, cells were incubated in fresh
349 media with Oregon Green 488-conjugated gelatin (10 µg/ml final concentration; Life
350 Technologies) for 1 hour at 37°C. Cells were then harvested (for suspension cells) or trypsinized
351 and harvested for analysis. To assess only internalized fluorescent collagen (i.e. to ignore
352 bound but not internalized collagen), cells were either washed before flow cytometry or Trypan
353 blue was added to quench bound but not internalized fluorochromes as we have done
354 previously[15,40]. For cell-surface staining, cells were stained using a standard FACS staining
355 protocol. Antibodies included: anti-MRC2 (as above) as a primary antibody with a secondary of
356 Donkey anti-sheep conjugated to a fluorophore (Invitrogen). Cells were then analyzed and/or
357 sorted on one of the above devices, with sorting and analysis gates based on fluorescence-
358 minus-one controls after compensation was performed with single-stain controls. For the

359 collagen uptake assay. Data were analyzed with FlowJo, version 7.6.1 (Tree Star, Ashland,
360 OR).
361
362 **Immunofluorescence Imaging**
363 Cells plated on coverslips were fixed with 4% PFA in PBS. Coverslips were blocked in
364 Phosphate-buffered saline (PBS) with 0.5% Bovine serum albumin, 0.1% Triton X-100 and 5%
365 donkey serum. Antibodies against Type I collagen (Southern Biotech, as above) or MRC2
366 (R&D, as above) or SEL1L (Novus, as above for Fig. 6A; for Fig. 6D, anti-SEL1L was provided
367 as a gift by Dr. Qi[28,47–49] and was used since it can recognize SEL1L without its FN2 domain)
368 was applied at 1:100 dilution followed by secondary antibody conjugated to Alexa Fluor 488 or
369 594 (Invitrogen) at 1:100. Coverslips were mounted in Vectashield (Vector Laboratories).
370 Human lung fragments were fixed in 4% PFA in PBS and then dehydrated in sucrose and
371 embedded in OCT and frozen in chilled isopropanol on dry ice. Frozen sections (30 μm) were
372 prepared. Sections were blocked in Phosphate-buffered saline (PBS) with 0.5% Bovine serum
373 albumin, 0.1% Triton X-100 and 5% donkey serum. Antibody against SEL1L (Novus, as above)
374 was applied at 1:100 dilution followed by secondary antibody conjugated to Alexa Fluor 488 or
375 594 (Invitrogen) at 1:100. Sections were mounted in Vectashield (Vector Laboratories).
376 Confocal images were captured using a Leica TCS SPE confocal microscope with an ACS APO
377 40x or 63x Oil CS objective lens at room temperature with Leica Type F Immersion Liquid
378 (11513859) or with a Leica Stellaris 8 confocal microscope with a 40x Oil objective. Wide-field
379 epifluorescence images were captured by an Olympus DP70 CCD using a standard Olympus
380 BX51 upright microscope with an Olympus UplanFL 40x objective at room temperature or with a
381 Nikon TE2000-U microscope and Ds-Qi2 camera. Images were processed with ImageJ
382 (National Institutes of Health, Bethesda, Maryland, USA). All imaging was done on the same
383 platform within any given experiment. Images were processed with ImageJ, in which any

384 brightness/contrast was changed uniformly across all conditions and pseudocoloring was
385 applied. Single optical slices at Airy 1 are shown for confocal images.
386
387 **Cell culture and molecular biology experiments**
388 U937 and MRC5 cells were obtained from ATCC. HEK293T cells that were WT, or null for
389 SEL1L or HRD1 were obtained from Dr. Qi and were described previously[28,47–49]. Cells were
390 grown in DMEM or RPMI (Corning) with 10% FBS (Gemini Bio-Products). ORFs containing full
391 length murine SEL1L or GAPDH were obtained from Origene and were maintained in pCMV6
392 expression vectors. FN2 deletion was accomplished using Q5 Site-Directed Mutagenesis kit
393 (New England Biolabs) according to the manufacturer's directions and in-frame deletion was
394 verified with Sanger sequencing. ORFs were also subcloned into pHIV lentiviral transfer
395 plasmids using standard techniques and lentiviral particles were packaged in HEK293T cells
396 with pMD2G and psPAX2. For MRC2 overexpression, pCI-Mrc2 (murine Mrc2 cloned into the
397 pCI-Neo expression vector, from Promega) was used as we previously described[13]. All
398 transfections were done using Lipofectamine 2000 (Invitrogen) according to the manufacturer's
399 instructions. For shRNA experiments, lentiviral particles harboring shRNA machinery within
400 pSicoR-puromycin vectors were obtained from Sigma-Aldrich (scramble control: SHC002V;
401 shRNA sequences can be found in Supplemental Table 2). Lentiviral infection was done in
402 these experiments at 10 multiplicity of infection (MOI) in cells overnight in complete medium
403 after addition of polybrene, and medium was replaced with fresh complete medium the next day.
404 The infected cells were allowed to grow and then selected by resistance to puromycin (1 μg/mll
405 Sigma-Aldrich). Puromycin-selected cells were collected for experiments as indicated. Cells
406 were infected at passages <10, and cells were collected for experiments between passages 5
407 and 16. For the luciferase assay, a 625bp fragment around the MRC2 transcriptional start site in
408 a Firefly luciferase vector with a minimal promoter element (Promega) was used as previously
409 described[13]. The Stop&Glo® (Promega) system was used to read luciferase activity in cells on a

410 luminometer (Synergy H1). All firefly luciferase values were normalized to luciferase levels of a
411 constitutive Renilla for a transfection control.
412
413 **Human lung tissue and primary human lung fibroblast isolation**
414 Normal human lung tissue from deceased donors of different ages was obtained from lungs not
415 used by the Northern California Transplant Donor Network according to an IRB-exempted
416 protocol led by Dr. Wolters; IPF lung specimens were obtained from explanted lungs removed
417 during lung transplantation, in an IRB-approved protocol led by Dr. Wolters. After harvest, lung
418 tissue was directly snap-frozen in liquid nitrogen and protein was extracted as above.
419
420 **Statistics**
421 Data were evaluated with Graphpad Prism software (San Diego, CA): for two sample
422 comparisons, by 2-tailed Student's *t*-test or Mann-Whitney *U* test if data did not pass a test of
423 normality; for multiple comparisons, by ANOVA with post-hoc Bonferroni testing if the ANOVA
424 was significant; or by Pearson correlation as indicated. A p-value of 0.05 or less was considered
425 statistically significant with *=p<0.05, **=p<0.01, ***=p<0.001, ****=p<0.0001, NS=Not
426 Significant. Data are presented as mean ± SEM.
427
428 **Study approval**
429 All experiments using human tissue were approved by the Weill Cornell Medicine or UCSF
430 Committee on Human Research. Informed consent was obtained for study participants or
431 waived for human tissue from deceased donors in accordance with the UCSF Committee on
432 Human Research.
433
434
435 Acknowledgments

This research was supported by US National Institutes of Health grants K08 HL145015 (MJP) and R01 HL136377 (KA), as well as a Sponsored Research Agreement with Celgene (KA). We acknowledge the UCSF Parnassus Flow CoLab (RRID:SCR_018206) for assistant generating flow cytometry data, supported in part by Grant NIH P30 DK063720 and by the NIH S10 Instrumentation Grant S10 1S10OD021822-01. We would like to thank S. Layer for ongoing inspiration.

Figure Legends

Figure 1. CRISPR screens identify mediators of cellular collagen uptake. (A) Schematic of experimental approach for flow cytometry-based screening for uptake of collagen fragments. (B) Volcano plot of gene-level statistics of CRISPRi (inhibition) screen. Most significant gene hits are labeled and colored in red. RRA = Robust Rank Aggregation score; LFC = Log-fold change. (C) Volcano plot of gene-level statistics of CRISPRa (activation) screen. (D) Comparison of overall phenotype score between CRISPRi and CRISPRa screens.

Figure 2. Collagen-biosynthetic pathways are regulators of collagen uptake uniquely compared to screens for uptake of other cargoes. (A) Ranked plot of Normalized Enrichment Scores derived from Gene-Set Enrichment Analysis (GSEA) of CRISPRi phenotype scores, using 10,561 Gene Ontology gene sets. Collagen-related gene sets are labeled and displayed in red. (B) Comparison of GSEA statistics for labeled collagen-related gene sets across screens for uptake of various cargoes[19]; screen method is indicated in parentheses. KD = Knock-down.

461 Figure 3. Collagen biosynthesis positively regulates cell-based collagen clearance and *MRC2*
462 expression. (A) Heatmap of overall phenotype scores from original screens as indicated for top
463 genes from 'Collagen Biosynthetic Process' GO gene-set. (B) Flow-cytometry based collagen-
464 uptake assay in U937 cells after shRNA treatment as indicated vs. Scramble control. MFI =
465 Mean Fluorescence Intensity. N=4 per group. (C) Q-RT-PCR in U937 cells after shRNA
466 treatment as indicated vs. Scramble control. N=4 per group. (D) Western blot of U937 cells after
467 shRNA treatment as indicated vs. Scramble control, representative of N=4 independent
468 experiments. (E) Q-RT-PCR in U937 cells after shRNA treatment as indicated vs. Scramble
469 control. N=4 per group. (F) Western blot of MRC5 fibroblasts after shRNA treatment as
470 indicated vs. Scramble control, representative of N=3 independent experiments. (G)
471 Representative confocal immunofluorescence images of MRC5 fibroblasts after shRNA
472 treatment as indicated vs. Scramble control. (H) Flow-cytometry based collagen-uptake assay in
473 MRC5 cells after shRNA treatment as indicated vs. Scramble control. N=4 per group. (I) Flow-
474 cytometry based collagen-uptake assay in U937 cells after shRNA treatment as indicated vs.
475 Scramble control. N=4-6 per group. (J-K) Q-RT-PCR in U937 cells after shRNA treatment as
476 indicated vs. Scramble control. N=4-5 per group. (L-M) Flow-cytometry based collagen-uptake
477 assay in U937 (L) or MRC5 (M) cells after drug treatment as indicated vs. vehicle control. N=4
478 per group. Western blot of corresponding cells and conditions shown below, each
479 representative of N=3-4 independent experiments. (N-P) Q-RT-PCR in MRC5 cells after drug
480 treatment as indicated vs. vehicle control. N=3-5 per group. Western blot of corresponding cells
481 and conditions shown below, each representative of N=3 independent experiments. Statistics:
482 (B-C, E, H-K) ANOVA with post-hoc Bonferroni testing; (L-P) Student's *t*-test. * $p < 0.05$, ** $p <$
483 0.01, **** $p < 0.0001$.
484
485 Figure 4. *Sel1L* positively regulates collagen uptake and *Mrc2* expression. (A) Schematic of
486 domain structure of SEL1L protein. FN2 = Fibronectin-2. (B) Flow-cytometry based collagen-

487 uptake assay in MEF cells (mouse embryonic fibroblasts) after treatment with 4-
488 hydroxytamoxifen (OHT) as indicated vs. vehicle control. WT = *Sel1L*$^{f/f;}$*ERcre*+, iKO =
489 *Sel1L*$^{f/f}$*;ERcre-*. N=3-4 per group. (C) Western blot of MEFs after OHT treatment as indicated vs.
490 control, representative of N=4 independent experiments. (D) Q-RT-PCR in MEFs after OHT
491 treatment for genes as indicated. N=3 per group. (E) Flow-cytometry based measurement of
492 MRC2 surface expression in MEFs after OHT treatment. N=4 per group. (F) Representative
493 wide-field immunofluorescence images of MEFs after OHT treatment. (G) Q-RT-PCR of MRC5
494 cells transfected with overexpression vectors containing genes as indicated vs. vehicle control
495 (lipofectamine only). N=4 per group. (H) Western blot of MRC5 cells transfected with
496 overexpression vectors containing genes as indicated vs. vehicle control (lipofectamine only).
497 N=2. Statistics: (B, G) ANOVA with post-hoc Bonferroni testing; (D-E) Student's *t*-test. ** $p <$
498 0.01, *** $p < 0.001$, **** $p < 0.0001$, ns = not significant.
499
500 Figure 5. SEL1L-mediated regulation of MRC2 is ERAD-independent. (A) Flow-cytometry based
501 collagen-uptake assay in 293T of different genotypes (compared with baseline control without
502 fluorescent collagen). N=4-5 per group. (B) Western blot of 293T cells of different genotypes,
503 representative of N=4 independent experiments. (C) Firefly luciferase assay of *MRC2* promoter
504 activity, normalized to constitutive Renilla luciferase in 293T cells of different genotypes. N=3
505 per group. (D) Western blot of HEK293T cells of different genotypes transfected with MRC2
506 overexpression vector vs. vehicle control, representative of N=2 independent experiments. (E)
507 Representative wide-field immunofluorescence images of 293T cells of different genotypes
508 transfected with MRC2 overexpression vector vs. vehicle control. (F) Flow-cytometry based
509 collagen-uptake assay in HEK293T of different genotypes transfected with MRC2
510 overexpression vector (compared with baseline control without fluorescent collagen; note
511 different scale compared with panel A). N=3-4 per group. Statistics: (A, C, F) ANOVA with post-
512 hoc Bonferroni testing. * $p < 0.05$ ** $p < 0.01$, ns = not significant.

513

514 Figure 6. SEL1L binds collagen as it is being synthesized via its FN2 domain and this interaction
515 is necessary for regulation of *MRC2* expression. (A) Representative confocal
516 immunofluorescence images of MRC5 fibroblasts treated as indicated or vehicle control. (B)
517 Western blot of MRC5 fibroblasts after immunoprecipitation with anti-SEL1L antibody (vs.
518 controls as indicated), representative of N=2 independent experiments. (C) Western blot after
519 immunoprecipitation with anti-procollagen type I antibody (with input controls as indicated) of
520 HEK293T cells transfected with collagen and WT or mutant SEL1L constructs as indicated,
521 representative of N=3 independent experiments. (D) Representative confocal
522 immunofluorescence images of HEK293T cells transfected with collagen and WT or mutant
523 SEL1L constructs as indicated. (E) Western blot of MEFs after treatment with WT or mutant
524 SEL1L constructs as indicated, representative of N=3 independent experiments. (F) Flow-
525 cytometry based measurement of MRC2 surface expression on MEFs after treatment with WT
526 or mutant SEL1L constructs as indicated, vs. Empty Vector (EV). N=6 per group. (G) Q-RT-PCR
527 of SEL1L-KO HEK293T cells after transfection with WT or mutant SEL1L as indicated vs. empty
528 vehicle control. N=3 per group. (H) Comparison of protein domain structures of Sel1L orthologs
529 across phylogenetically diverse organisms; FN2 = Fibronectin-2 domain, tpr = Tetratricopeptide
530 repeat domain. Statistics: (F-G) ANOVA with post-hoc Bonferroni testing. * $p < 0.05$, ** $p < 0.01$,
531 ns = not significant.

532

533 Figure 7. Collagen biosynthesis and *MRC2* expression are positively correlated in lung tissue,
534 but this relationship is impaired in human fibrotic lung disease. (A) UMAP plot with cell type
535 annotation of scRNAseq data from multiple available datasets of human lung control and IPF
536 tissue (GSE136831, GSE135893, GSE121611, GSE128033, GSE132771). N=184,672 single
537 cells. (B-C) Gene expression of individual genes as indicated in UMAP as generated in panel
538 (A). (D) Comparison of gene expression in all cells from data as in panel (A) that had non-zero

539 expression of *COL1A1* or *MRC2*. (E) Comparison of gene expression in whole human lung
540 microarray data from GSE110147 analyzed via GEO2R in pulmonary fibrosis lungs vs. control
541 specimens. N=11 vs. 37 per group. (F) Comparison of microarray gene expression in human
542 lung bronchoalveolar lavage specimens from GSE70867 analyzed via GEO2R from Idiopathic
543 Pulmonary Fibrosis (IPF) lungs vs. control specimens. N=20 vs. 112. (G) Comparison of
544 microarray gene expression in peripheral blood mononuclear cell (PBMC) specimens from
545 GSE37858 analyzed via GEO2R from Idiopathic Pulmonary Fibrosis (IPF) lungs vs. control
546 specimens. N=35 vs. 70. (H) Comparison of gene expression in whole mouse lung microarray
547 data from GSE40151 analyzed via GEO2R from Bleomycin-treated mice vs. Saline control
548 mice. N=55 vs. 56. Statistics: ANCOVA, p-values indicated in figures. Pearson correlations are
549 displayed in panel (D).
550
551 Figure 8. SEL1L protein levels are reduced in the lungs of patients with Idiopathic Pulmonary
552 Fibrosis. (A-B) Western blot of whole human lung lysates from IPF lungs vs. control (donor
553 lungs deemed not suitable for transplant), with corresponding densitometry. N=7 per group. (C)
554 Representative low-power wide-field immunofluorescence images of human lungs (IPF vs.
555 control). (D) Q-RT-PCR of whole human lungs (IPF vs. control; same specimens as used in
556 panels A and B). Statistics: (B, D) Mann-Whitney *U* test. * $p < 0.05$, ns = not significant.
557
558
559 Supplemental Figures
560
561 Supplemental Figure 1. Re-screens are correlated to original CRISPR screen results. (A)
562 Volcano plot of CRISPRi re-screen (Screen 2). (B) Volcano plot of CRISPRa re-screen (Screen
563 2). (C) Comparison of gene-level overall phenotype scores between CRISPRi screens 1 and 2.
564 (D) Comparison of gene-level overall phenotype scores between CRISPRa screens 1 and 2. (E)

565 Comparison of arrayed log-fold change in arrayed validation vs. original pooled screen log-fold
566 change; each point represents unique guide sequence. 65/84 guides demonstrated
567 concordance with original pooled screen data (77%). Pearson correlation coefficient is shown.
568
569 Supplemental Figure 2. Collagen-related gene ontology terms are enriched as regulators of
570 collagen uptake in CRISPRi and CRISPRa screens. Top twenty positive and negatively
571 performing gene sets are indicated with associated statistics for each screen as indicated,
572 among all 10,561 Gene Ontology gene sets. Collagen-related gene sets are highlighted with red
573 boxes. NES = Normalized Enrichment Score.
574
575
576 Supplemental Tables
577
578 Supplemental Table 1. Gene-level data from CRISPRi and CRISPRa screens of collagen
579 uptake as analyzed via MAGeCK.
580
581 Supplemental Table 2. Primer and shRNA sequences used in this manuscript.
582
583
584 References
585
586 1. McKleroy, W., Lee, T.-H. & Atabai, K. Always cleave up your mess: targeting collagen
587 degradation to treat tissue fibrosis. *AJP: Lung Cellular and Molecular Physiology* **304**,
588 L709–L721 (2013).

589 2. Atabai, K. *et al.* Mfge8 diminishes the severity of tissue fibrosis in mice by binding and
590 targeting collagen for uptake by macrophages. *Journal of Clinical Investigation* (2009)
591 doi:10.1172/JCI40053.
592 3. Bundesmann, M. M. *et al.* Role of urokinase plasminogen activator receptor-associated
593 protein in mouse lung. *Am J Respir Cell Mol Biol* **46**, 233–9 (2012).
594 4. Montaño, M. *et al.* Lung collagenase inhibitors and spontaneous and latent collagenase
595 activity in idiopathic pulmonary fibrosis and hypersensitivity pneumonitis. *Chest* **96**, 1115–
596 9 (1989).
597 5. Selman, M. *et al.* TIMP-1, -2, -3, and -4 in idiopathic pulmonary fibrosis. A prevailing
598 nondegradative lung microenvironment? *Am J Physiol Lung Cell Mol Physiol* **279**, L562-
599 74 (2000).
600 6. Montfort, I. & Pérez-Tamayo, R. Collagenase in experimental carbon tetrachloride
601 cirrhosis of the liver. *Am J Pathol* **92**, 411–20 (1978).
602 7. Okazaki, I. & Maruyama, K. Collagenase activity in experimental hepatic fibrosis. *Nature*
603 **252**, 49–50 (1974).
604 8. Brady, A. H. Collagenase in scleroderma. *J Clin Invest* **56**, 1175–80 (1975).
605 9. Engelholm, L. H. *et al.* uPARAP/Endo180 is essential for cellular uptake of collagen and
606 promotes fibroblast collagen adhesion. *J Cell Biol* **160**, 1009–15 (2003).
607 10. Yuan, C. *et al.* Crystal structures of the ligand-binding region of uPARAP: effect of
608 calcium ion binding. *Biochem J* **473**, 2359–68 (2016).
609 11. Madsen, D. H. *et al.* Endocytic collagen degradation: A novel mechanism involved in
610 protection against liver fibrosis. *Journal of Pathology* **227**, 94–105 (2012).
611 12. López-Guisa, J. M. *et al.* Mannose receptor 2 attenuates renal fibrosis. *J Am Soc Nephrol*
612 **23**, 236–51 (2012).
613 13. Podolsky, M. J. *et al.* Age-dependent regulation of cell-mediated collagen turnover. *JCI*
614 *Insight* **5**, (2020).

615 14. Melander, M. C., Jürgensen, H. J., Madsen, D. H., Engelholm, L. H. & Behrendt, N. The
616 collagen receptor uPARAP/Endo180 in tissue degradation and cancer (Review).
617 *International Journal of Oncology* vol. 47 1177–1188 Preprint at
618 https://doi.org/10.3892/ijo.2015.3120 (2015).
619 15. Lee, T.-H. *et al.* Functional genomic screen identifies novel mediators of collagen uptake.
620 **25**, (2014).
621 16. Messaritou, G., East, L., Roghi, C., Isacke, C. M. & Yarwood, H. Membrane type-1 matrix
622 metalloproteinase activity is regulated by the endocytic collagen receptor Endo180. *J Cell*
623 *Sci* **122**, 4042–4048 (2009).
624 17. Gilbert, L. A. *et al.* Genome-Scale CRISPR-Mediated Control of Gene Repression and
625 Activation. *Cell* **159**, 647–661 (2014).
626 18. Madsen, D. H. *et al.* The non-phagocytic route of collagen uptake: A distinct degradation
627 pathway. *Journal of Biological Chemistry* (2011) doi:10.1074/jbc.M110.208033.
628 19. Haney, M. S. *et al.* Identification of phagocytosis regulators using magnetic genome-wide
629 CRISPR screens. *Nat Genet* **50**, 1716–1727 (2018).
630 20. Stefanovic, B., Stefanovic, L., Schnabl, B., Bataller, R. & Brenner, D. A. TRAM2 protein
631 interacts with endoplasmic reticulum Ca2+ pump Serca2b and is necessary for collagen
632 type I synthesis. *Mol Cell Biol* **24**, 1758–68 (2004).
633 21. Bucala, R., Spiegel, L. A., Chesney, J., Hogan, M. & Cerami, A. Circulating fibrocytes
634 define a new leukocyte subpopulation that mediates tissue repair. *Mol Med* **1**, 71–81
635 (1994).
636 22. Phillips, R. J. *et al.* Circulating fibrocytes traffic to the lungs in response to CXCL12 and
637 mediate fibrosis. *J Clin Invest* **114**, 438–46 (2004).
638 23. Kleaveland, K. R. *et al.* Fibrocytes are not an essential source of type I collagen during
639 lung fibrosis. *J Immunol* **193**, 5229–39 (2014).

640 24. Simões, F. C. *et al.* Macrophages directly contribute collagen to scar formation during
641 zebrafish heart regeneration and mouse heart repair. *Nat Commun* **11**, 600 (2020).
642 25. Ozhogina, O. A., Trexler, M., Bányai, L., Llinás, M. & Patthy, L. Origin of fibronectin type
643 II (FN2) modules: structural analyses of distantly-related members of the kringle family
644 idey the kringle domain of neurotrypsin as a potential link between FN2 domains and
645 kringles. *Protein Sci* **10**, 2114–22 (2001).
646 26. Bányai, L. & Patthy, L. Evidence for the involvement of type II domains in collagen
647 binding by 72 kDa type IV procollagenase. *FEBS Lett* **282**, 23–5 (1991).
648 27. Pickford, A. R., Potts, J. R., Bright, J. R., Phan, I. & Campbell, I. D. Solution structure of a
649 type 2 module from fibronectin: implications for the structure and function of the gelatin-
650 binding domain. *Structure* **5**, 359–370 (1997).
651 28. Sun, S. *et al.* Sel1L is indispensable for mammalian endoplasmic reticulum-associated
652 degradation, endoplasmic reticulum homeostasis, and survival. *Proceedings of the*
653 *National Academy of Sciences* **111**, (2014).
654 29. Christianson, J. C. *et al.* Defining human ERAD networks through an integrative mapping
655 strategy. *Nat Cell Biol* **14**, 93–105 (2012).
656 30. Mueller, B., Lilley, B. N. & Ploegh, H. L. SEL1L, the homologue of yeast Hrd3p, is
657 involved in protein dislocation from the mammalian ER. *J Cell Biol* **175**, 261–70 (2006).
658 31. Biunno, I. *et al.* Cross-species conservation of SEL1L, a human pancreas-specific
659 expressing gene. *OMICS* **6**, 187–98 (2002).
660 32. Fidler, A. L. *et al.* A unique covalent bond in basement membrane is a primordial
661 innovation for tissue evolution. *Proc Natl Acad Sci U S A* **111**, 331–6 (2014).
662 33. Hecker, L. *et al.* Reversal of persistent fibrosis in aging by targeting Nox4-Nrf2 redox
663 imbalance. *Sci Transl Med* **6**, (2014).
664 34. Peng, R. *et al.* Bleomycin induces molecular changes directly relevant to idiopathic
665 pulmonary fibrosis: a model for ‘active’ disease. *PLoS One* **8**, e59348 (2013).

666 35. Morse, C. *et al.* Proliferating SPP1/MERTK-expressing macrophages in idiopathic
667 pulmonary fibrosis. *Eur Respir J* **54**, (2019).
668 36. Reyfman, P. A. *et al.* Single-Cell Transcriptomic Analysis of Human Lung Provides
669 Insights into the Pathobiology of Pulmonary Fibrosis. *Am J Respir Crit Care Med* **199**,
670 1517–1536 (2019).
671 37. Habermann, A. C. *et al.* Single-cell RNA sequencing reveals profibrotic roles of distinct
672 epithelial and mesenchymal lineages in pulmonary fibrosis. *Sci Adv* **6**, eaba1972 (2020).
673 38. Adams, T. S. *et al.* Single-cell RNA-seq reveals ectopic and aberrant lung-resident cell
674 populations in idiopathic pulmonary fibrosis. *Sci Adv* **6**, eaba1983 (2020).
675 39. Tsukui, T. *et al.* Collagen-producing lung cell atlas identifies multiple subsets with distinct
676 localization and relevance to fibrosis. *Nat Commun* **11**, 1920 (2020).
677 40. Podolsky, M. J. *et al.* Cell division cycle 7 kinase is a negative regulator of cell-mediated
678 collagen degradation. *Am J Physiol Lung Cell Mol Physiol* **315**, L360–L370 (2018).
679 41. Boettcher, M. *et al.* Dual gene activation and knockout screen reveals directional
680 dependencies in genetic networks. *Nat Biotechnol* **36**, 170–178 (2018).
681 42. Boettcher, M. *et al.* Tracing cellular heterogeneity in pooled genetic screens via multi-
682 level barcoding. *BMC Genomics* **20**, 107 (2019).
683 43. Li, W. *et al.* MAGeCK enables robust identification of essential genes from genome-scale
684 CRISPR/Cas9 knockout screens. *Genome Biol* **15**, 554 (2014).
685 44. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for
686 interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545–50
687 (2005).
688 45. Mootha, V. K. *et al.* PGC-1alpha-responsive genes involved in oxidative phosphorylation
689 are coordinately downregulated in human diabetes. *Nat Genet* **34**, 267–73 (2003).
690 46. Altenhoff, A. M. *et al.* OMA orthology in 2021: website overhaul, conserved isoforms,
691 ancestral gene order and more. *Nucleic Acids Res* **49**, D373–D379 (2021).

692 47. Shrestha, N. *et al.* Sel1L-Hrd1 ER-associated degradation maintains β cell identity via
693 TGF-β signaling. *J Clin Invest* **130**, 3499–3510 (2020).
694 48. Zhou, Z. *et al.* Endoplasmic reticulum–associated degradation regulates mitochondrial
695 dynamics in brown adipocytes. *Science (1979)* **368**, 54–60 (2020).
696 49. Bhattacharya, A. *et al.* Hepatic Sel1L-Hrd1 ER-associated degradation (ERAD) manages
697 FGF21 levels and systemic metabolism via CREBH. *EMBO J* **37**, (2018).
698

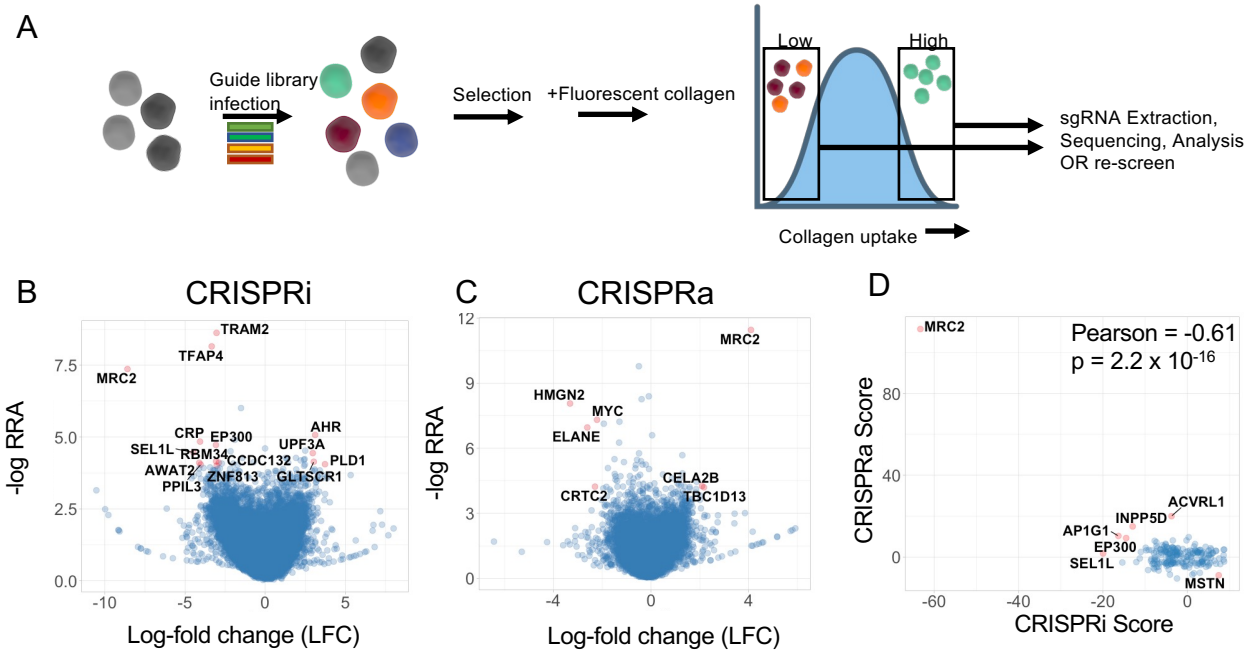
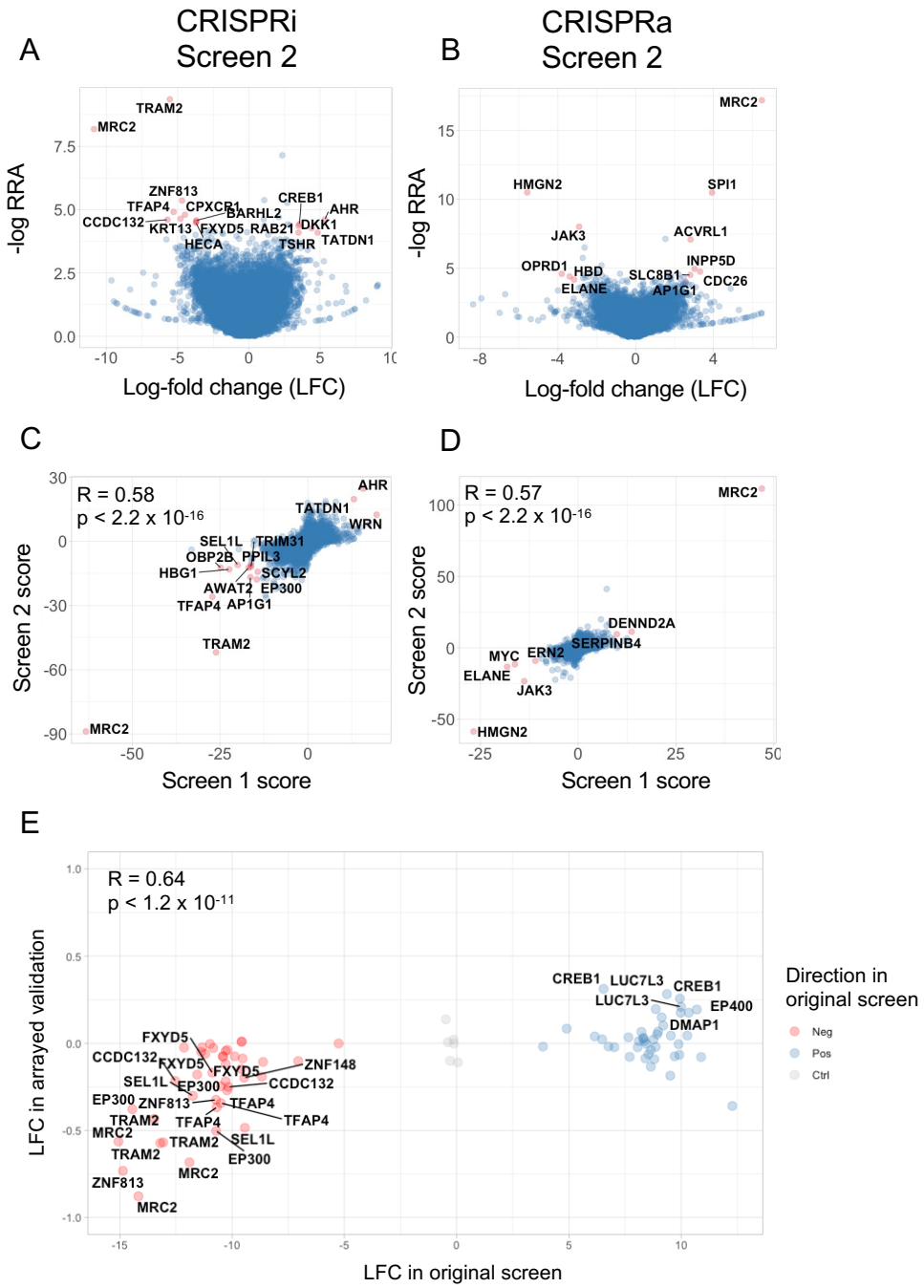


Figure 1. CRISPR screens identify mediators of cellular collagen uptake. (A) Schematic of experimental approach for flow cytometry-based screening for uptake of collagen fragments. (B) Volcano plot of gene-level statistics of CRISPRi (inhibition) screen. Most significant gene hits are labeled and colored in red. RRA = Robust Rank Aggregation score; LFC = Log-fold change. (C) Volcano plot of gene-level statistics of CRISPRa (activation) screen. (D) Comparison of overall phenotype score between CRISPRi and CRISPRa screens.



Supplemental Figure 1. Re-screens are correlated to original CRISPR screen results. (A) Volcano plot of CRISPRi re-screen (Screen 2). (B) Volcano plot of CRISPRa re-screen (Screen 2). (C) Comparison of gene-level overall phenotype scores between CRISPRi screens 1 and 2. (D) Comparison of gene-level overall phenotype scores between CRISPRa screens 1 and 2. (E) Comparison of arrayed log-fold change in arrayed validation vs. original pooled screen log-fold change; each point represents unique guide sequence. 65/84 guides demonstrated concordance with original pooled screen data (77%). Pearson correlation coefficient is shown.

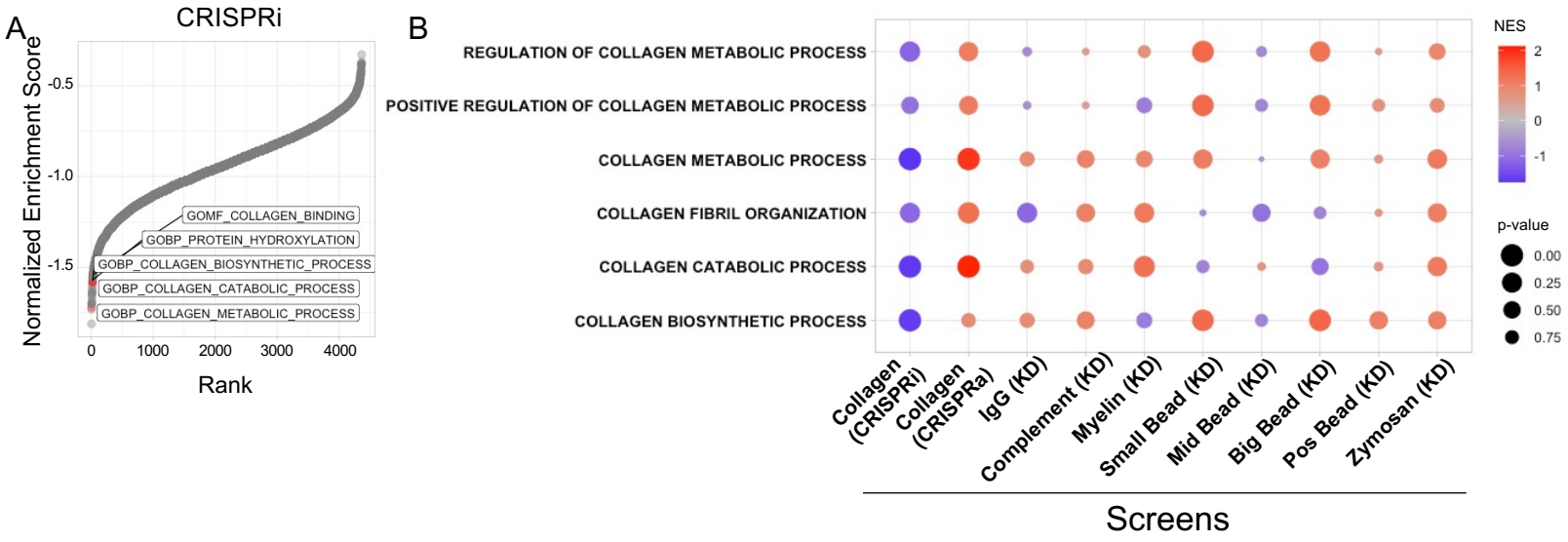
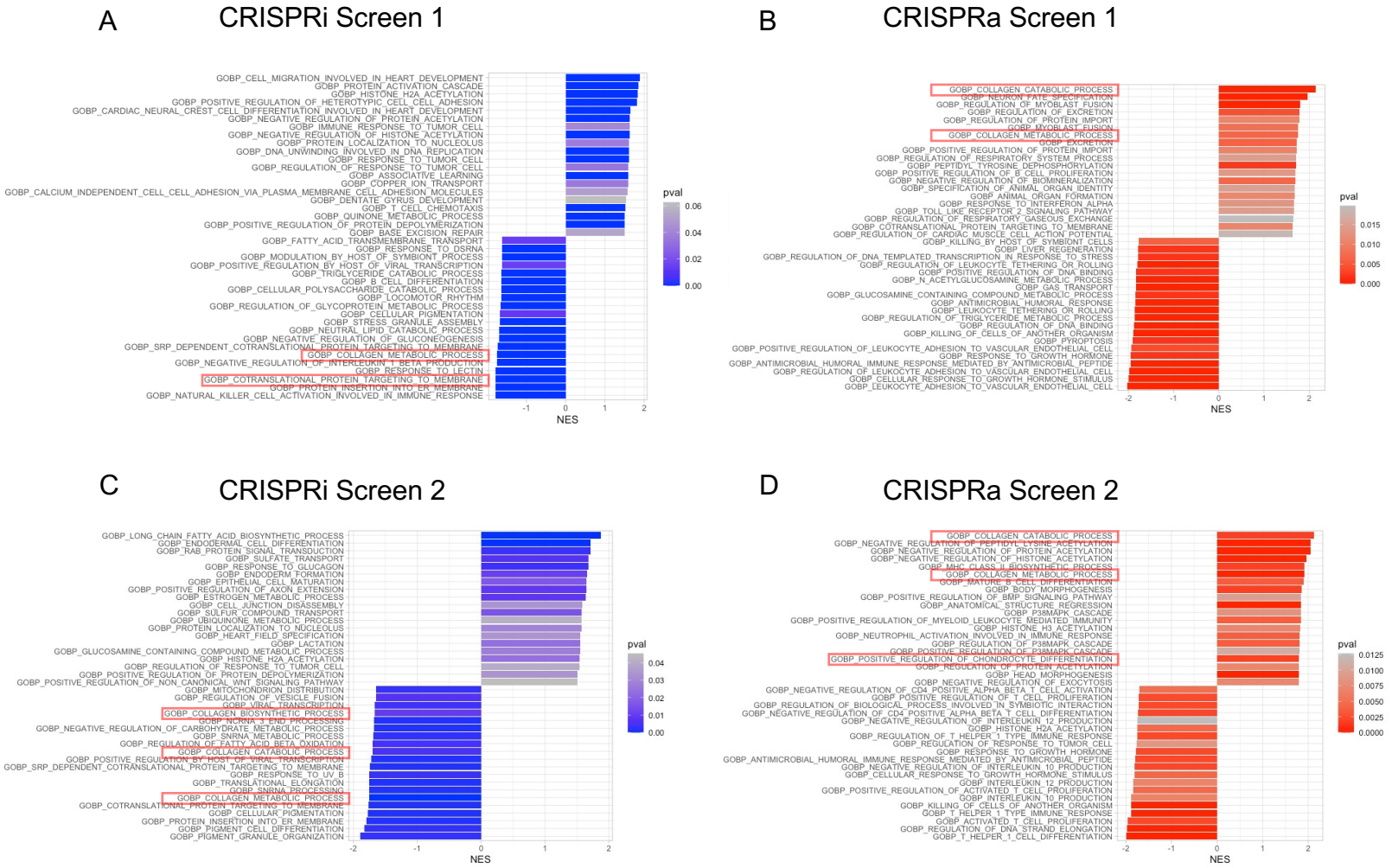


Figure 2. Collagen-biosynthetic pathways are regulators of collagen uptake uniquely compared to screens for uptake of other cargoes. (A) Ranked plot of Normalized Enrichment Scores derived from Gene-Set Enrichment Analysis (GSEA) of CRISPRi phenotype scores, using 10,561 Gene Ontology gene sets. Collagen-related gene sets are labeled and displayed in red. (B) Comparison of GSEA statistics for labeled collagen-related gene sets across screens for uptake of various cargoes[19]; screen method is indicated in parentheses. KD = Knock-down.



Supplemental Figure 2. Collagen-related gene ontology terms are enriched as regulators of collagen uptake in CRISPRi and CRISPRa screens. Top twenty positive and negatively performing gene sets are indicated with associated statistics for each screen as indicated, among all 10,561 Gene Ontology gene sets. Collagen-related gene sets are highlighted with red boxes. NES = Normalized Enrichment Score.

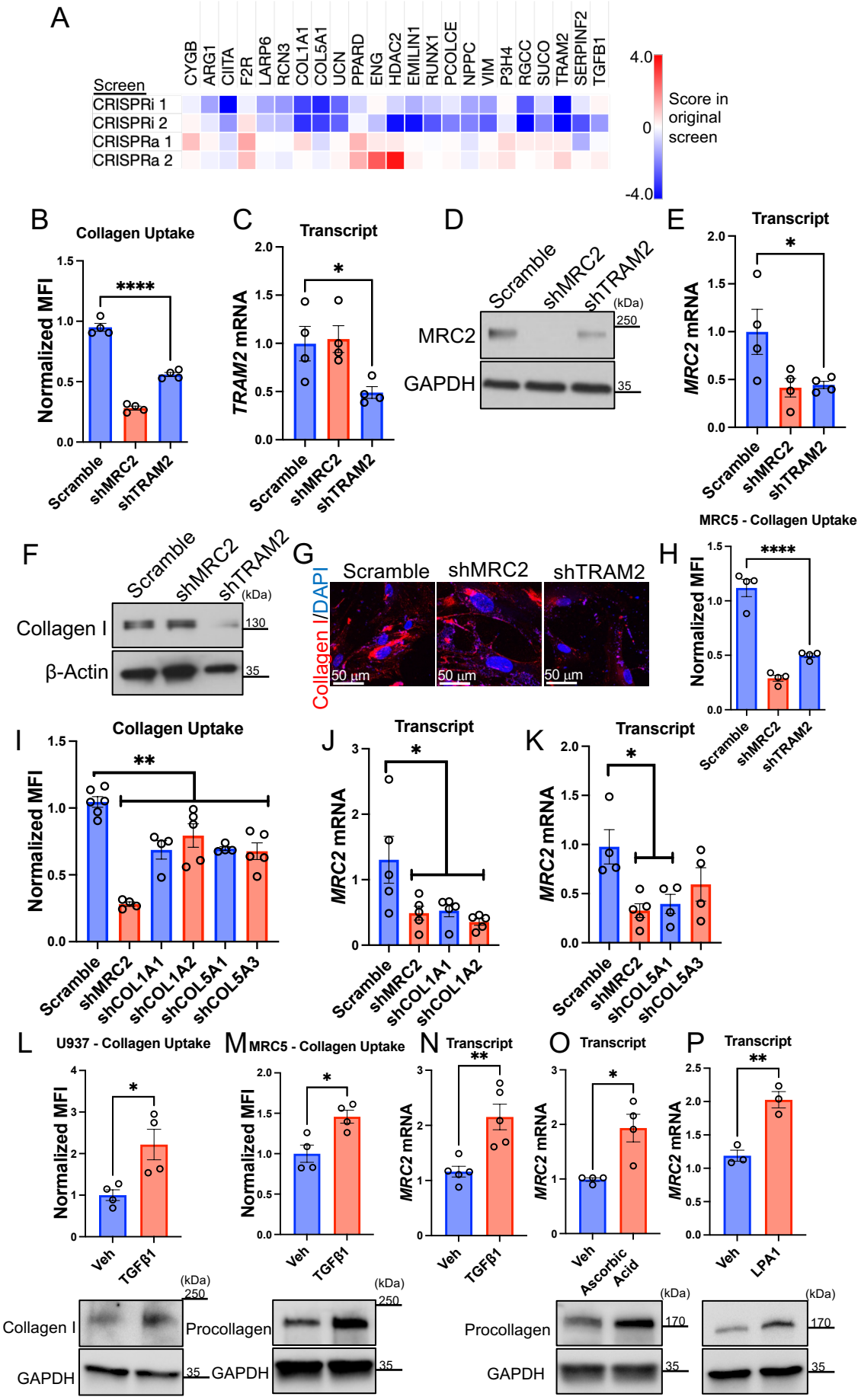


Figure 3. Collagen biosynthesis positively regulates cell-based collagen clearance and *MRC2* expression. (A) Heatmap of overall phenotype scores from original screens as indicated for top genes from 'Collagen Biosynthetic Process' GO gene-set. (B) Flow-cytometry based collagen-uptake assay in U937 cells after shRNA treatment as indicated vs. Scramble control. MFI = Mean Fluorescence Intensity. N=4 per group. (C) Q-RT-PCR in U937 cells after shRNA treatment as indicated vs. Scramble control. N=4 per group. (D) Western blot of U937 cells after shRNA treatment as indicated vs. Scramble control, representative of N=4 independent experiments. (E) Q-RT-PCR in U937 cells after shRNA treatment as indicated vs. Scramble control. N=4 per group. (F) Western blot of MRC5 fibroblasts after shRNA treatment as indicated vs. Scramble control, representative of N=3 independent experiments. (G) Representative confocal immunofluorescence images of MRC5 fibroblasts after shRNA treatment as indicated vs. Scramble control. (H) Flow-cytometry based collagen-uptake assay in MRC5 cells after shRNA treatment as indicated vs. Scramble control. N=4 per group. (I) Flow-cytometry based collagen-uptake assay in U937 cells after shRNA treatment as indicated vs. Scramble control. N=4-6 per group. (J-K) Q-RT-PCR in U937 cells after shRNA treatment as indicated vs. Scramble control. N=4-5 per group. (L-M) Flow-cytometry based collagen-uptake assay in U937 (L) or MRC5 (M) cells after drug treatment as indicated vs. vehicle control. N=4 per group. Western blot of corresponding cells and conditions shown below, each representative of N=3-4 independent experiments. (N-P) Q-RT-PCR in MRC5 cells after drug treatment as indicated vs. vehicle control. N=3-5 per group. Western blot of corresponding cells and conditions shown below, each representative of N=3 independent experiments. Statistics: (B-C, E, H-K) ANOVA with post-hoc Bonferroni testing; (L-P) Student's *t*-test. * $p < 0.05$, ** $p < 0.01$, **** $p < 0.0001$.

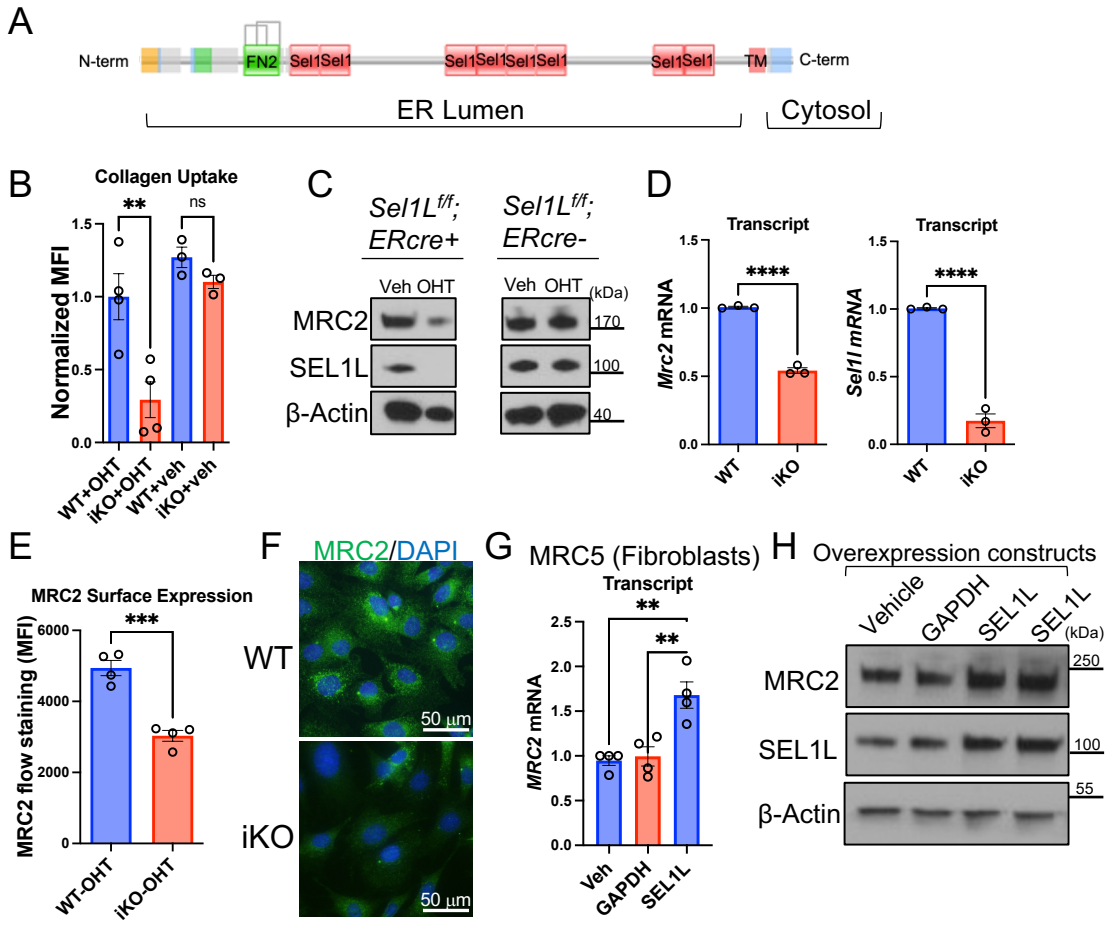


Figure 4. *Sel1L* positively regulates collagen uptake and *Mrc2* expression. (A) Schematic of domain structure of SEL1L protein. FN2 = Fibronectin-2. (B) Flow-cytometry based collagen-uptake assay in MEF cells (mouse embryonic fibroblasts) after treatment with 4-hydroxytamoxifen (OHT) as indicated vs. vehicle control. WT = *Sel1L$^{f/f}$;ERcre*+, iKO = *Sel1L$^{f/f}$;ERcre*-. N=3-4 per group. (C) Western blot of MEFs after OHT treatment as indicated vs. control, representative of N=4 independent experiments. (D) Q-RT-PCR in MEFs after OHT treatment for genes as indicated. N=3 per group. (E) Flow-cytometry based measurement of MRC2 surface expression in MEFs after OHT treatment. N=4 per group. (F) Representative wide-field immunofluorescence images of MEFs after OHT treatment. (G) Q-RT-PCR of MRC5 cells transfected with overexpression vectors containing genes as indicated vs. vehicle control (lipofectamine only). N=4 per group. (H) Western blot of MRC5 cells transfected with overexpression vectors containing genes as indicated vs. vehicle control (lipofectamine only). N=2. Statistics: (B, G) ANOVA with post-hoc Bonferroni testing; (D-E) Student's *t*-test. ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$, ns = not significant.

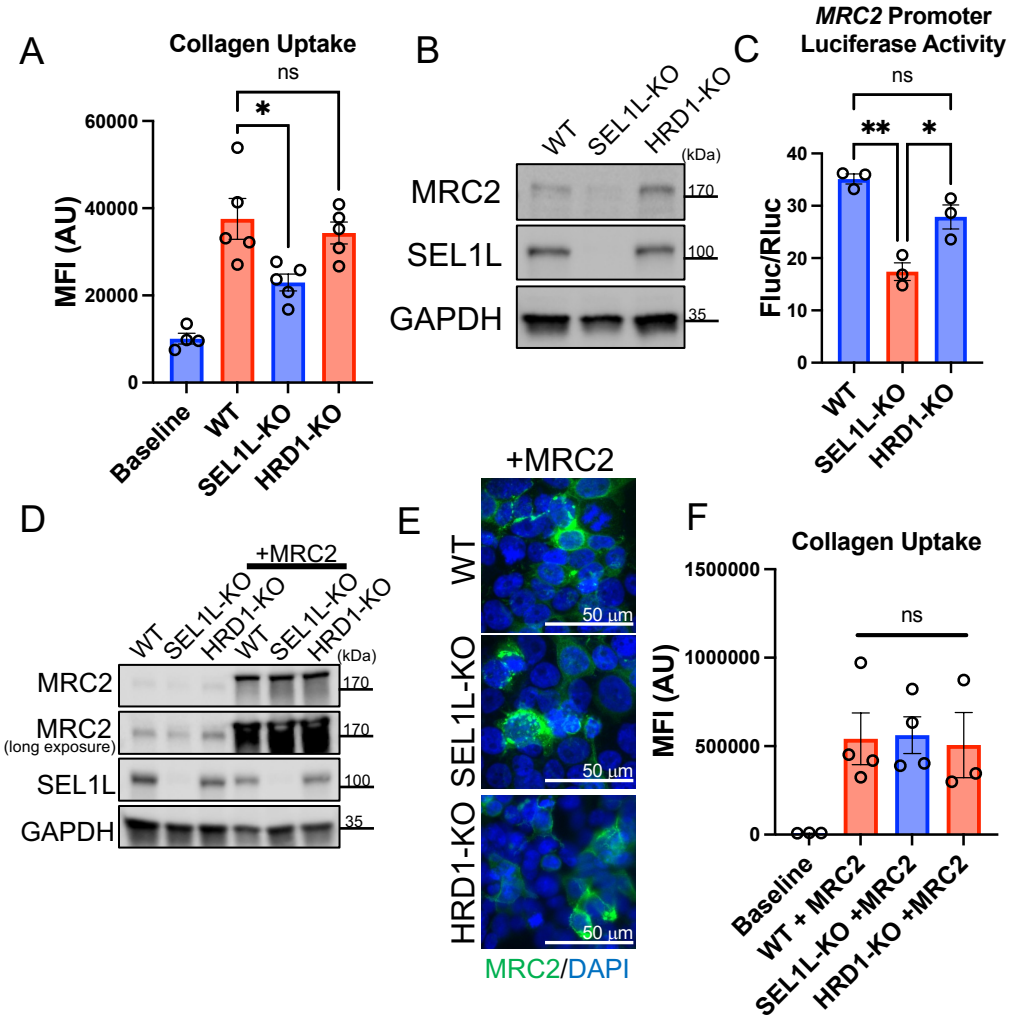


Figure 5. SEL1L-mediated regulation of MRC2 is ERAD-independent. (A) Flow-cytometry based collagen-uptake assay in 293T of different genotypes (compared with baseline control without fluorescent collagen). N=4-5 per group. (B) Western blot of 293T cells of different genotypes, representative of N=4 independent experiments. (C) Firefly luciferase assay of *MRC2* promoter activity, normalized to constitutive Renilla luciferase in 293T cells of different genotypes. N=3 per group. (D) Western blot of HEK293T cells of different genotypes transfected with MRC2 overexpression vector vs. vehicle control, representative of N=2 independent experiments. (E) Representative wide-field immunofluorescence images of 293T cells of different genotypes transfected with MRC2 overexpression vector vs. vehicle control. (F) Flow-cytometry based collagen-uptake assay in HEK293T of different genotypes transfected with MRC2 overexpression vector (compared with baseline control without fluorescent collagen; note different scale compared with panel A). N=3-4 per group. Statistics: (A, C, F) ANOVA with post-hoc Bonferroni testing. * $p < 0.05$ ** $p < 0.01$, ns = not significant.

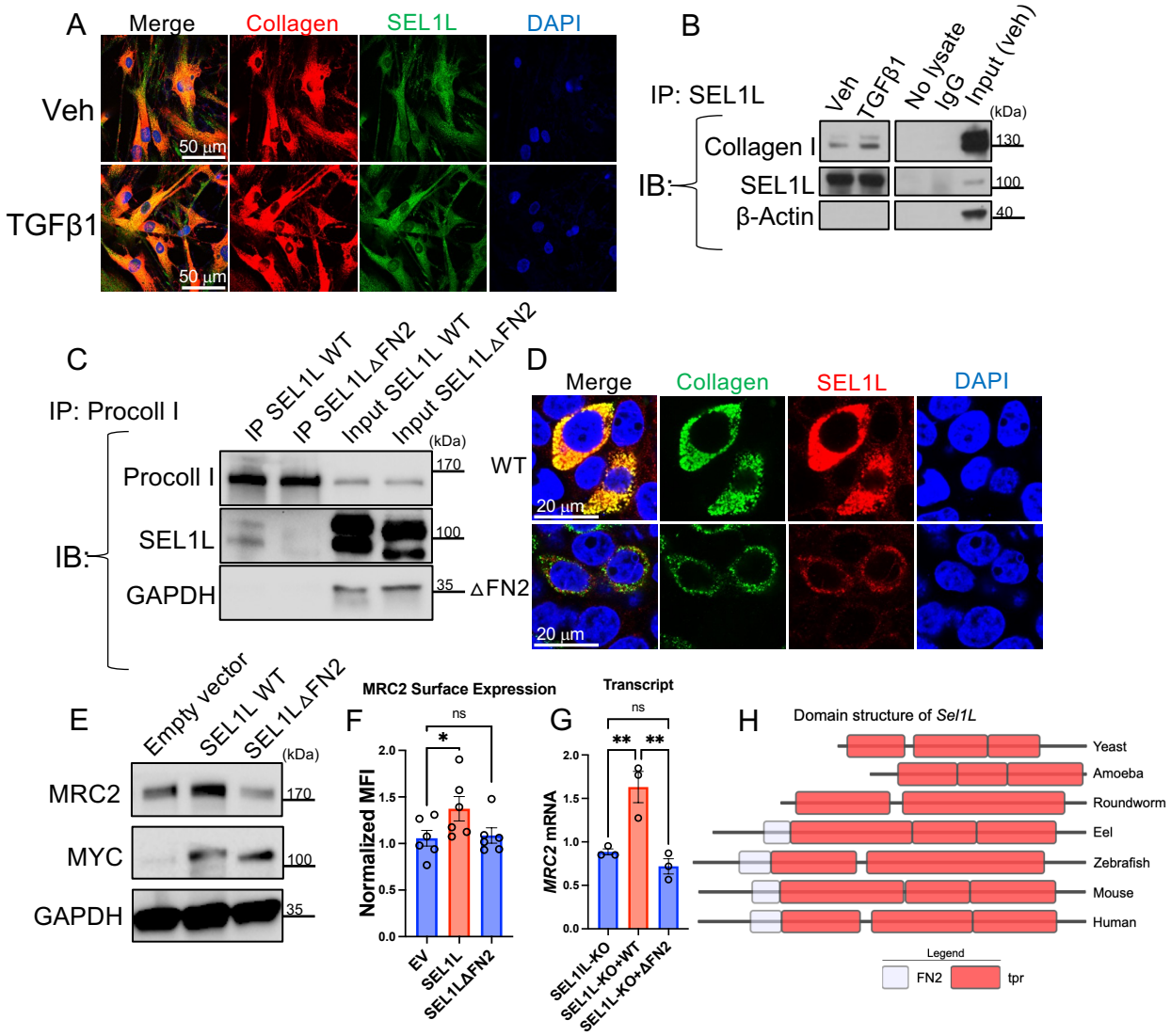


Figure 6. SEL1L binds collagen as it is being synthesized via its FN2 domain and this interaction is necessary for regulation of *MRC2* expression. (A) Representative confocal immunofluorescence images of MRC5 fibroblasts treated as indicated or vehicle control. (B) Western blot of MRC5 fibroblasts after immunoprecipitation with anti-SEL1L antibody (vs. controls as indicated), representative of N=2 independent experiments. (C) Western blot after immunoprecipitation with anti-procollagen type I antibody (with input controls as indicated) of HEK293T cells transfected with collagen and WT or mutant SEL1L constructs as indicated, representative of N=3 independent experiments. (D) Representative confocal immunofluorescence images of HEK293T cells transfected with collagen and WT or mutant SEL1L constructs as indicated. (E) Western blot of MEFs after treatment with WT or mutant SEL1L constructs as indicated, representative of N=3 independent experiments. (F) Flow-cytometry based measurement of MRC2 surface expression on MEFs after treatment with WT or mutant SEL1L constructs as indicated, vs. Empty Vector (EV). N=6 per group. (G) Q-RT-PCR of SEL1L-KO HEK293T cells after transfection with WT or mutant SEL1L as indicated vs. empty vehicle control. (H) Comparison of protein domain structures of Sel1L orthologs across phylogenetically diverse organisms; FN2 = Fibronectin-2 domain, tpr = Tetratricopeptide repeat domain. N=3 per group. Statistics: (F-G) ANOVA with post-hoc Bonferroni testing. * $p < 0.05$, ** $p < 0.01$, ns = not significant.

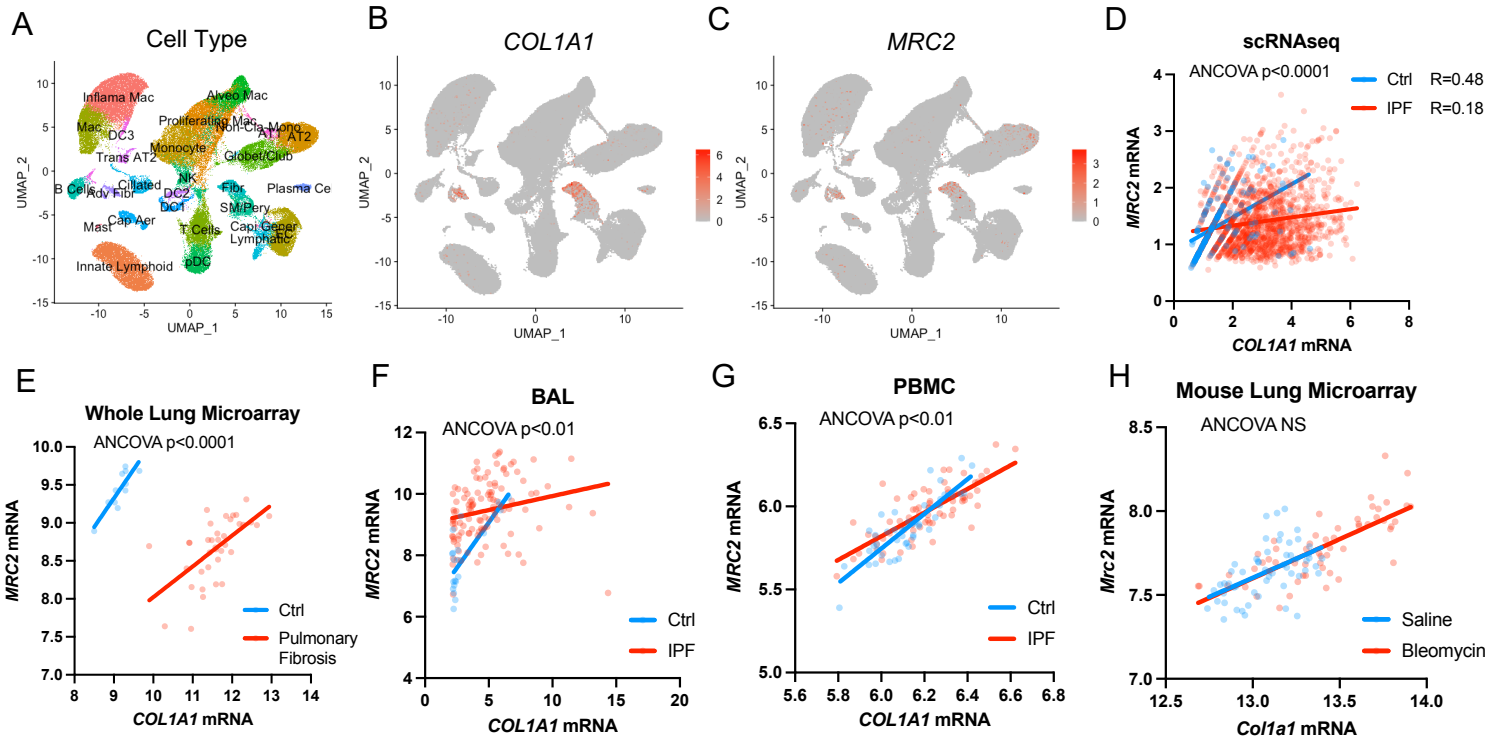


Figure 7. Collagen biosynthesis and *MRC2* expression are positively correlated in lung tissue, but this relationship is impaired in human fibrotic lung disease. (A) UMAP plot with cell type annotation of scRNAseq data from multiple available datasets of human lung control and IPF tissue (GSE136831, GSE135893, GSE121611, GSE128033, GSE132771). N=184,672 single cells. (B-C) Gene expression of individual genes as indicated in UMAP as generated in panel (A). (D) Comparison of gene expression in all cells from data as in panel (A) that had non-zero expression of *COL1A1* or *MRC2*. (E) Comparison of gene expression in whole human lung microarray data from GSE110147 analyzed via GEO2R in pulmonary fibrosis lungs vs. control specimens. N=11 vs. 37 per group. (F) Comparison of microarray gene expression in human lung bronchoalveolar lavage specimens from GSE70867 analyzed via GEO2R from Idiopathic Pulmonary Fibrosis (IPF) lungs vs. control specimens. N=20 vs. 112. (G) Comparison of microarray gene expression in peripheral blood mononuclear cell (PBMC) specimens from GSE37858 analyzed via GEO2R from Idiopathic Pulmonary Fibrosis (IPF) lungs vs. control specimens. N=35 vs. 70. (H) Comparison of gene expression in whole mouse lung microarray data from GSE40151 analyzed via GEO2R from Bleomycin-treated mice vs. Saline control mice. N=55 vs. 56. Statistics: ANCOVA, p-values indicated in figures. Pearson correlations are displayed in panel (D).

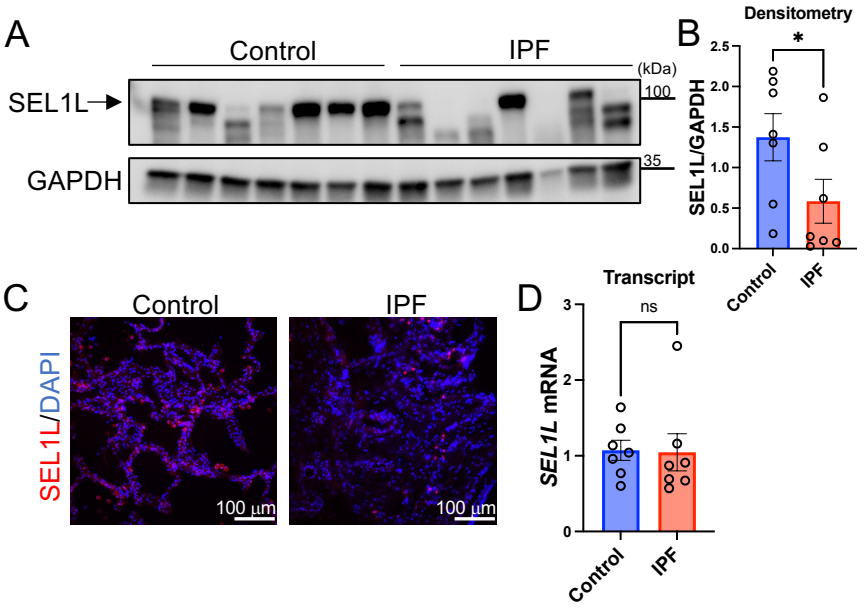


Figure 8. SEL1L protein levels are reduced in the lungs of patients with Idiopathic Pulmonary Fibrosis. (A-B) Western blot of whole human lung lysates from IPF lungs vs. control (donor lungs deemed not suitable for transplant), with corresponding densitometry. N=7 per group. (C) Representative low-power wide-field immunofluorescence images of human lungs (IPF vs. control). (D) Q-RT-PCR of whole human lungs (IPF vs. control; same specimens as used in panels A and B). Statistics: (B, D) Mann-Whitney *U* test. * $p < 0.05$, ns = not significant.