UCSF

UC San Francisco Previously Published Works

Title

Extracellular Matrix Orchestration of Tissue Remodeling in the Chronically Inflamed Mouse Colon

Permalink

https://escholarship.org/uc/item/2nj2p2cd

Journal

Cellular and Molecular Gastroenterology and Hepatology, 17(4)

ISSN

2352-345X

Authors

Moutin, Elisa B

Bons, Joanna

Giavara, Giada

et al.

Publication Date

2024

DOI

10.1016/j.jcmgh.2024.01.003

Peer reviewed

Extracellular matrix orchestration of tissue remodelling in the chronically inflamed mouse colon

Short title: ECM drivers of tissue remodelling in colitis

Elisa B. Moutin[1], Joanna Bons[2], Giada Giavara[1], Filipe Lourenco[1], Deng Pan[3], Jordan B. Burton[2], Samah Shah[2], Mathilde Colombé[1], Philippe Gascard[3], Thea Tlsty[3], Birgit Schilling[2] and Douglas J Winton[1]

[1]*Cancer Research-UK Cambridge Institute, Li Ka Shing Centre, Robinson Way, Cambridge, CB2 0RE, UK.*

[2]*Buck Institute for Research on Aging, Novato, California, USA.*

[3]*Department of Pathology, University of California, San Francisco, California, USA.*

Correspondence

Dr. Douglas J. Winton

Cancer Research-UK Cambridge Institute, Li Ka Shing Centre, Robinson Way, Cambridge, CB2 0RE, UK.

Email: doug.winton@cruk.cam.ac.uk

Ph: +44 (0) 1223 769783

Conflict of interest statement

The authors disclose no conflicts.

Data Availability Statement

All data relevant to this study are available from the corresponding authors upon reasonable request. Raw data and complete MS data sets have been uploaded to the MassIVE repository of the Center for Computational Mass Spectrometry at UCSD, and can be downloaded using the   following   link: https://massive.ucsd.edu/ProteoSAFe/dataset.jsp?task=a172756fa86c4d64ab8331e58554a dd0 (MassIVE ID number: MSV000092056; ProteomeXchange ID: PXD042551).

[Note to the reviewers: To access the data repository MassIVE (UCSD) for MS data, please use: Username: MSV000092056_reviewer; Password: winter].

[Note to the reviewers: To access the data repository MassIVE (UCSD) for MS data, please use: Username: xx_reviewer; Password: winter].

Abbreviations: ECM, extracellular matrix; IBD, inflammatory bowel disease; UC, ulcerative colitis; DIA, data-independent-acquisition; SPF, specific-pathogen-free; LC-MS/MS, liquid chromatography with tandem mass spectrometry; GO, gene ontology; FDR, false discovery rate; SLRPs, small leucine rich proteoglycans; PCA, principal component analysis; CIACs, chronic-inflammation-associated cancers.

56 **Abstract**

57 **Background & Aims:** Chronic inflammatory illnesses are debilitating and recurrent conditions

58 associated with significant co-morbidities, including an increased risk of developing cancer.

59 Extensive tissue remodelling is a hallmark of such illnesses, and is both a consequence and

60 a mediator of disease progression. Despite previous characterisation of epithelial and stromal

61 remodelling during Inflammatory Bowel Disease, a complete understanding of its impact on

62 disease progression is still lacking.

63 **Methods:** A comprehensive proteomic pipeline using data-independent acquisition was

64 applied to decellularized colon samples from the $Muc2^{KO}$ mouse model of colitis, for an in-

65 depth characterisation of extracellular matrix remodelling. Unique proteomic profiles of the

66 matrisomal landscape were extracted from pre- and pathological colitis. Integration of

67 proteomics and transcriptomics datasets extracted from the same murine model produced

68 network maps describing the orchestrating role of matrisomal proteins in tissue remodelling

69 during the progression of colitis.

70 **Results:** The in-depth proteomic workflow used here allowed the addition of 34 proteins to

71 the known colon matrisomal signature. Protein signatures of pre- and pathological colitic

72 states were extracted, differentiating the two states by expression of small leucine rich

73 proteoglycans. We outlined the role of this class and other matrisomal proteins in tissue

74 remodelling during colitis, as well as the potential for co-ordinated regulation of cell types by

75 matrisomal ligands.

76 **Conclusion:** Our work highlights a central role for matrisomal proteins in tissue remodelling

77 during colitis and defines orchestrating nodes that can be exploited in selection of therapeutic

78 targets.

79 **Keywords:** Inflammatory Bowel Disease, DIA Proteomics, Cell-matrix interactions,

80 Proteoglycans

81

82

Introduction

Chronic inflammation generates compositional and adaptive changes at the cellular and molecular level. Resulting tissue remodelling ultimately impairs epithelial function, leading to diverse disease states such as chronic liver disease, gastritis or colitis, which are debilitating and life altering conditions associated with an increased risk of developing cancer[1]. Characterisation of these chronic illnesses has largely been cell-centric, describing epithelial and stromal cell adaptations while viewing the extracellular matrix (ECM) and its associated proteins - the matrisome - as a supporting scaffold.

The matrisome is comprised of core ECM constituents (collagens, glycoproteins, proteoglycans and polysaccharides) as well as associated regulators and secreted factors[2]. It is produced by resident cell types that are found in unique combinations in different tissues[3]. Evidence for an instructive capacity of the ECM originally came from studies where separation of embryonic epithelium and mesenchyme before heterotypic recombination revealed that the mesenchyme was responsible for tissue differentiation[4]. Subsequently, many studies have established that the matrisome is dynamically maintained and associated with varied cellular responses, including cell growth and differentiation through direct ligand-receptor interactions[5]. Furthermore, it dictates biomechanical properties and is able to sequester and regulate the availability of cytokines and growth factors[6]. More recently, embryonic precursor cells seeded in complete ECM scaffolds derived from decellularized tissues were found to differentiate into epithelia of the same tissue type as that from which the scaffold was derived[7-9].

The relationship whereby the matrisome dynamically regulates the biology of the cells from which itself arises has been termed 'dynamic reciprocity'[10]. During chronic inflammation, excessive degradation of the ECM by proteases prevents matrix composition changes necessary for wound healing. This process termed fibrosis is linked with increased cancer risk[11,12]. The ECM has been shown to influence all the hallmarks that typify development of

4

111 cancer, and its remodelling is considered central to cancer initiation due to chronic

112 inflammation[13,15].

113

114 Ulcerative Colitis (UC) is a chronic inflammatory disorder prone to remission and relapse, with

115 a complex and not yet entirely understood aetiology involving both genetic and environmental

116 factors[16]. UC is defined by colonic mucosal inflammation associated with reduced quality and

117 quantity of the mucins that make up the intestinal mucus layer, impaired epithelial barrier

118 function and engagement with both innate and adaptive immune systems[17,18]. UC patients are

119 at a greater risk of developing colitis-associated colorectal cancer, which increases with

120 duration and severity of disease, reaching 18% by 30 years[19,20].

121

122 The complexity of tissue response and disease escalation in UC, as well as technological

123 barriers in quantifying protein changes in fractions with low solubility/low abundance have

124 meant that the full repertoire of matrisomal changes as well as how and when these

125 orchestrate cellular adaptation and recruitment in UC have not been appreciated. Recently,

126 label-free mass-spectrometry-based proteomic strategies based on comprehensive data-

127 independent acquisition (DIA)[21-23] have been employed to identify proteins in ECM-enriched

128 cancer tissue samples and provided deep, accurate and reproducible quantification[24]. Here

129 we apply this unbiased approach to a murine model of colitis to characterise alterations in

130 matrisomal protein expression. Network analyses define how ECM changes converge on key

131 signalling nodes in both pre- and pathological states to identify vulnerabilities for therapeutic

132 and preventive interventions.

133

134 **Results**

135 **Experimental approach to investigate ECM remodelling in colitis**

136 Germline genetic knockout of the *Muc2* gene in mice has been previously described as

137 causing colitis that arises from impaired barrier function[25, 26]. This model was phenotypically

138 assessed and mice with homozygous deletion of Mucin2 (*Muc2*[hom]) were confirmed to show

139   stochastic onset of colitis (Supplemental Figure 1A, B). Staining with both Alcian blue and

140   Pas revealed a reduction in expression of both acid and neutral mucins, the remaining staining

141   showing that goblet cells are retained (Supplemental Figure 1A). This mouse model faithfully

142   recapitulated key features of human UC including crypt enlargement and hyperplasia, immune

143   infiltration of the mucosa, and presence of abscesses (Supplemental Figure 1B), as well as

144   pre-neoplastic features such as dysplasia and early invasive epithelial foci (Supplemental

145   Figure 1C). In mature *Muc2^{hom}* mice (5-8 months old), these phenotypes were present

146   regionally in both proximal and distal colon, whilst the middle colon remained histologically

147   normal (Supplemental Figure 1D, E).

148

149   Samples for proteomic analysis were collected from five mature *Muc2^{hom}* mice aged until they

150   presented with diarrhoea, symptom of active colitis, and culled when this was accompanied

151   with other symptoms such as signs of pain (hunch, piloerection) and/or development of an

152   anal prolapse. Tissue from five age-matched mice *Muc2^{het}* mice were concurrently used as

153   controls (Supplemental Figure 2A). The colon tissue was taken from proximal, actively

154   inflamed regions and from middle uninvolved regions (Supplemental Figure 2B). Samples

155   were processed for enrichment of ECM proteins by sequential fractionation that was validated

156   at each stage by Western blotting (Supplemental Figure 3A). The final protein fraction was

157   solubilised, digested and subjected to mass-spectrometry analysis as described previously[24]

158   (Supplemental Figure 3B). A spectral library was directly built from DIA MS/MS spectra and

159   used to identify and quantify all proteins across all conditions. In total 2,328 protein groups

160   were identified and quantified, with at least two unique peptides and a 1% FDR. A clustering

161   analysis based on all quantified protein groups led to removal of 2 outlier samples that

162   clustered apart from their own sample group (Supplemental Figure 3C). For the remaining

163   samples, correlation coefficients for protein expression within sample group was above 85%

164   (Supplemental Figure 3D). Principal Component Analysis revealed genotype and location

165   along the colon as the main factors contributing to variance between sample groups

166   (Supplemental Figure 3E). Of the 2,328 proteins identified, 953 (~41% of the total protein

6

167 fraction) were annotated as 'extracellular' using the Gene Ontology term "cellular component"

168 (Figure 1A). Within this subset, 125 proteins were identified as components or associated

169 proteins of the colon matrisome (MatrisomeDB[2.0])[27], accounting for 70% of all annotated

170 proteins in this dataset. In addition, 37 identified proteins were part of the wider matrisome

171 database compiling data from different organs, that can now be added to the known colon

172 matrisomal signature (Supplemental Table S3).

173

174 **Proteomic analysis reflects disease state**

175 Proteins with expression significantly altered in colitis (defined as Qvalue < 0.05), identified

176 from the comparison of proximal colonic regions of *Muc2^hom^* inflamed and *Muc2^het^* control mice,

177 were assessed for alterations in biological processes, using the molecular interaction

178 database ConsensusPathDB. The analysis identified positive enrichment for ECM

179 organisation as well as different immune-related functions (Figure 1B, C), confirming that the

180 decellularization protocol had enriched for ECM components and that the resulting dataset

181 was informative of ECM remodelling during colitis.

182

183 The analysis focused on proteins with low solubility, which as reported previously, include

184 cellular remnants such as keratins and desmosomal proteins that relate to overall tissue and

185 cellular identity[28]. To evaluate how protein alterations reflect cell population level changes

186 accompanying inflammatory disease, differentially expressed proteins (defined as Qvalue <

187 0.05) from the comparison of *Muc2^hom^* and *Muc2^het^* proximal colons were assigned to specific

188 cell types, using previously described signatures[30-32] (see Methods and Supplemental Figure

189 3F). This analysis identified enrichment for immune cell types and under-representation of

190 epithelial secretory cells, notably goblet cells, reflecting well known features of colitis (Figure

191 1D, E). These findings indicated that proteomic analysis of the extracellular protein fraction

192 captured the complexity of tissue changes associated with disease state at the level of both

193 biological processes and cell type identity.

194

Deriving a colitis signature

To derive a robust colitis signature, we selected the top 200 proteins with the highest significance for differential expression in the proximal colon of *Muc2hom* versus location matched *Muc2het* samples (Supplemental Table S4). Many upregulated proteins were associated with specific immune cell types (e.g., eosinophils, Epx; neutrophils, Mpo; and mast cells, Mcpt) or immune function (Igha, H2-D1, Tap1, Tap2, Irgm1) (Figure 1F). Included in this signature was also the downregulation of small leucine-rich proteoglycans (SLRPs) Asporin (Aspn) and Osteoglycin (Ogn), both present in the subset of proteins presented in Figure 1F, as well as Decorin (Dcn), present in the extended inflammation signature. SLRPs are known to bind and alter the properties of collagens[29]. At the level of protein groups, proteins classified as proteoglycans showed the most dramatic change in expression with the majority being downregulated with inflammation (Figure 1G).

Mid-colon in mature *Muc2hom* mice displays a pre-pathological state

Comparison of the mid-colonic regions from *Muc2hom* and *Muc2het* mice identified 1,712 differentially expressed proteins (Figure 2A). Of these, 721 were also differentially expressed in the proximal colonic region of *Muc2hom* versus *Muc2het* mice that showed signs of active colitis. (Figure 2A). The similarity between proximal inflamed and middle histologically normal colon samples was apparent in the Principal Component Analysis (PCA), where those samples overlapped (Supplemental Figure 3E).

Comparing differentially expressed proteins in proximal and middle colon samples from *Muc2hom* (compared to location matched *Muc2het* samples) revealed a linear relationship with a positive correlation, showing that shared alterations also follow the same direction of change (Supplemental Figure 4A). We hypothesized that the substantial overlap in the number of proteins sharing the same direction of change might be explained by two opposing interpretations. First, it may reflect stable adaptive changes accompanying loss of *Muc2* expression. Alternatively, it could indicate an early pre-pathological stage in the development

8

223   of inflammatory disease. To discriminate between these possibilities, a cohort of *Muc2^hom^* mice

224   were maintained with colitis symptoms (median survival 12 months, range: 11.7 to 12.6

225   months) and analysed for inflammation along the length of the colon (Supplemental Figure 2).

226   These older animals showed an overall increase in pathology-related metrics, as well as a

227   trend towards pan-colitis (Supplemental Figures 4B, C). This observation supported that the

228   middle region of the colon in mature *Muc2^hom^* mice constituted a pre-pathological state.

229

230   A shared representative signature was derived from proteins differentially expressed in the

231   comparison of inflamed and pre-pathological *Muc2^hom^* versus location matched *Muc2^het^*

232   samples and that shared the same direction of change (Figure 2A, B; Supplemental Table S4

233   "shared signature"). Sult1a1, a protein involved in small molecule metabolism and previously

234   found downregulated in DSS colitis[30] was downregulated in both pre-pathological and

235   inflamed states (Figure 2B, C). In contrast, Gremlin 1 (Grem1) expression increased

236   incrementally in the respective states (Figure 2B, D). Notably, immunohistochemistry analysis

237   revealed that while Grem1 expression was restricted to the basement membrane in the pre-

238   pathological state, it extended to the mucosa in inflamed tissues (Figure 2E).

239

240   As the mid colonic mucosa remained histologically normal in 5–8-month-old *Muc2^hom^* mice on

241   initial presentation, we speculated that the observed molecular changes may illustrate an

242   adaptation of the epithelium to inflammatory signals. Proteins were identified that were

243   differentially expressed and going in the opposite direction when comparing mid and proximal

244   inflamed regions of *Muc2^hom^* mice to their respective control regions in *Muc2^het^* samples

245   (Figure 2A, F; Supplemental Table S4 "opposite signature"). Reciprocating patterns of

246   expression were identified for ECM proteins including collagens (Col15a1, Col6a5, Col6a6)

247   and SLRP proteoglycans (Prelp, Dcn), as well as proteins implicated in mitochondrial function

248   (Vdac1, Vdac3, Ndufs, Acads) that were all upregulated in mid-colon and down regulated in

249   proximal inflamed colon (Figure 2F, H). Intriguingly, also exhibiting this pattern was Aoc3 (Vap-

250   1), a protein known to exist in tissue-bound and soluble forms and detected primarily in

endothelial cells, adipocytes or serum, respectively. The tissue-bound form has been associated with tissue differentiation and ECM deposition whereas the soluble form has been shown to be pro-inflammatory and implicated in vascular diseases through involvement in leukocyte recruitment[31]. Showing the opposing pattern, i.e., down-regulated in mid colon and up-regulated in inflamed proximal was Chemokine ligand 6 (Ccl6) (Figure 2F, G). Increased expression of this chemoattractant chemokine is consistent with the elevated immune cell recruitment observed in inflammation.

Differential expression of proteoglycans distinguishes pathological states

Expression of six small-leucine-rich proteoglycans (SLRPs) in mid and proximal inflamed regions of mature *Muc2[hom]* mice was assessed relative to their respective control regions in *Muc2[het]* samples. These canonical SLRPs[32] play diverse roles in collagen deposition, bind to receptor tyrosine kinases, innate immune receptors and modulate the TGF-β signalling pathway [33].

All identified SLRPs, in particular Prolargin (Prelp) and Decorin (Dcn), were downregulated with inflammation but not in mid-colon, illustrating a clear distinction in differential proteoglycan expression associated with early (pre-pathological) and advanced pathological states (Figure 3A, B). The Bader lab CellCellInteraction database[34] was used to explore the consequences of altered SLRP expression by identifying differentially expressed interactors in both states. Decorin (Dcn) and biglycan (Bgn) presented with the highest number of known interactors (Figure 3C). Pathway analysis based on differentially expressed SLRPs interactors (see Material and Methods) revealed a strong enrichment for inflammation-related pathways such as 'inflammatory response' and 'allograft rejection' in the inflamed proximal but not the mid colon of *Muc2[hom]* mice (Figure 3D). Interestingly, an enrichment for the p53 pathway was observed in the mid colon of *Muc2[hom]* mice, contrasting with an upregulation of the Kras pathway in inflamed proximal samples. Heightened p53 activity in mid-colon was accompanied with positive enrichment for apoptosis, suggesting a possible role in control of

279   cellular integrity. No interactors of SLRPs could be mapped to the p53 or apoptosis pathway

280   in proximal inflamed samples, suggesting that this pathway might be downregulated to allow

281   for cell proliferation necessary for tissue repair. Taken together, these findings highlight the

282   central role that downregulation of proteoglycans plays in the progression of chronic

283   inflammation.

284

285   **Matrisomal orchestration of tissue remodelling in colitis**

286   Next, to infer the orchestrating role of matrisomal proteins in tissue remodelling, the 162

287   matrisomal proteins previously identified (Supplemental Table S3, Figure 1A) were

288   interrogated for differential expression in mature *Muc2^{hom}* mice, selecting only proteins

289   annotated as ligands. The CellCellInteraction database[34] was then used to infer potential

290   interactions using receptors found to be differentially expressed in *Muc2^{hom}* mice at the

291   transcriptomic level. In inflamed proximal *Muc2^{hom}* tissue, 47 matrisomal proteins were

292   identified as ligands to 27 differentially expressed receptors. In contrast, in mid colon only 10

293   receptors and 37 matrisomal proteins were differentially expressed (Supplemental Figure 5A,

294   Supplemental Table S5), showing that the matrisomal regulome in pre-pathology maps to

295   significantly fewer interactors than in inflamed tissue. For example, Fibronectin (Fn1)

296   interacted with 25 receptors in inflamed tissue but with only 7 in pre-pathology (Supplemental

297   Figure 5B). Further, the repertoire of differentially expressed integrins is reduced in pre-

298   pathology. The only integrins engaging with matrisomal ligands were Itga6 and Itgb6,

299   compared to the inflamed state where Itgav, Itgae, Itgal, Itgax and Itgb2 were additionally

300   engaged. Taken together, this shows a greater complexity of regulation in overt inflammation

301   compared to pre-pathology.

302

303   Receptors interacting with matrisomal ligands were then mapped to cell types to reveal the

304   extent of the matrisomal contribution to regulation of tissue function in the context of colitis

305   (Supplemental Figure 5C). This analysis indicated that 10 cell types were potentially regulated

306   by differentially expressed matrisomal proteins in inflamed tissue, compared to 8 in pre-

11

307 pathology (Supplemental Figure 5D). In both states, endothelial cells had the most

308 connections to matrisomal proteins. Matrisomal regulation of macrophages and B cells

309 appeared limited to inflamed tissue (Figure 5A). Proteins such as Fibronectin were implicated

310 in engagement of 8 different cell types in inflammation. In pre-pathology, engagement of 4 cell

311 types was the maximum for any matrisomal protein (Supplemental Figure 5E).

312

313 Finally, an integrated network was created to link differentially expressed matrisomal proteins

314 to the cell types they regulate (Figure 4A). This network revealed greater involvement of ECM

315 regulators and secreted factors in inflammation compared to pre-pathology, and more

316 widespread tissue landscape alterations extending to include immune cell types such as B

317 cells, T cells and macrophages.

318

319 Hierarchical clustering was used to recognise similarities in the regulation of different cell types

320 by matrisomal proteins altered in inflammation (Figure 4B). This analysis revealed a possible

321 co-ordinated regulation of fibroblasts and endothelial cells by a subset of proteins, including

322 Lysyl Oxidase homolog 2 (Loxl2) and Neutrophil Elastase (Elane). Of note, matrisomal

323 proteins such as S100 calcium-binding protein A9 (s100a9), Plasminogen (Plg) and Loxl2

324 suggest a shared function in increasing collagen cross-linking (Figure 4B).

325

326 To gain direct evidence for this co-ordinated regulation by matrisomal ligands, the tissue

327 distribution of macrophages, fibroblasts and endothelial cells was first compared using

328 immunofluorescence in $Muc2^{hom}$ pre-pathological mid colon and inflamed proximal tissue

329 (Figure 4C, D). In the former, CD31+ endothelial cells, CD68+ macrophages and Serpinh1+

330 fibroblasts are primarily located in a region adjacent to the lower crypt epithelium and

331 submucosa. With inflammation, fibroblasts and endothelial cells show a marked displacement

332 towards the luminal surface at the top of the crypts, whilst macrophages largely remain in a

333 more basal location. Next, the expression profile of four matrisomal ligands (Lamc1, Anxa1,

334 S100a9 and Elane) that are predominantly associated with fibroblast and endothelial

335 regulation along the crypt axis was similarly assessed (Figure 4B, E). This revealed a trend

336 for increased expression of these ligands in the same luminal region with inflammation. Of

337 note, Lamc1 expression was also displaced towards the basal side of crypts (Figure 4F).

338 These findings suggest that spatial changes in the frequency of fibroblasts and endothelial

339 cells can be tracked to the ligands associated with their regulation. Together, this data defines

340 the orchestrating role of the matrisome in the development of colonic inflammatory disease.

341

342 **Discussion**

343 Despite widespread acceptance of the fundamental role of the matrisome in regulating tissue

344 homeostasis, and remodelling in chronic disease states such as colitis, relatively few studies

345 have focused on understanding and integrating the full repertoire of matrisomal

346 orchestration[35]. Here, the development of label-free mass spectrometry (DIA-MS) approaches

347 to quantify proteins in ECM enriched samples provided the opportunity to build an integrated

348 picture of matrisomal regulation in a mouse model that captures much of the pathology and

349 dysfunction seen in UC patients[25,26,36]. Importantly, the choice of model also enabled

350 identification of ECM remodelling events and candidate pathways involved in early tissue

351 adaptation to chronic inflammation prior to any histological manifestation.

352

353 Combining protein fractionation with comprehensive and sensitive data-independent-

354 acquisition mass-spectrometry revealed the extensive tissue remodelling accompanying

355 colitis and related it to mediators of the inflammatory phenotype. For example, within the

356 colitis signature latent TGF-β proteins (Ltbp) are enriched and are responsible for directing

357 latent TGF-β to extracellular matrix microfibrils where it becomes bioavailable upon tissue

358 remodelling to mediate both inflammatory and fibrotic responses seen both in the *Muc2*[KO]

359 model and in UC patients[37][38]. In addition, the depth of this unbiased proteomic analysis

360 allowed deconvolution of cell populations known to be associated with colonic inflammation

361 in both murine models of colitis as well as in the human disease, using approaches

362 previously applied at the transcriptomic level using single cell technologies[39-41].

363 Many differentially expressed proteins were detected in the apparently histologically normal

364 middle colon, which led to the finding that this tissue had already acquired phenotypes that

365 qualify it as occupying a pre-pathological state. This finding is reminiscent of recent

366 observations in a delayed onset model of induced colitis that identified protein changes prior

367 to the onset of inflammatory disease[35]. Analysis of colitis in older animals confirmed the

368 development of a pan-colitis, allowing interrogation of the proteomic data for adaptive

369 responses that may restrain histological manifestations of inflammation at earlier disease

370 stages.

371

372 One class of proteins that may be involved in restricting inflammatory disease are SLRPs, that

373 are elevated in pre-pathology prior to their reduction in overt inflammation. Downregulation of

374 small leucine-rich proteoglycans is associated with poor outcomes in invasive breast cancer[42],

375 and in chronic-inflammation-associated cancers (CIACs), including lung[24]. Interrogating

376 interactors of SLRPs differentially expressed in pre-pathology, we found evidence for

377 increased p53 activity that was not seen in the inflamed state. This is in accord with previous

378 studies showing that the proliferative phase of tissue regeneration requires the coordinated

379 upregulation of trophic pathways (as seen with Kras here) while downregulating that of p53[43].

380 Hence, the transient elevated expression of SLRPs in the pre-pathological state may coincide

381 with the activation of a p53 checkpoint control that is subsequently lost as inflammatory

382 disease develops, resulting in a corresponding increase in cancer risk.

383 Notably, the pre-pathological state was associated with elevated expression of many

384 mitochondrial proteins. In IBD patients, genetic risk factors affecting mitochondrial function

385 have been identified and their altered metabolism proposed as causative in inflammatory

386 disease[44-46]. Recently, altered mitochondrial function has been shown to predict disease

387 recurrence in Crohn's disease supporting a role for mitochondrial dysfunction developing prior

388 to the onset of active inflammation[47].

389

390   To uncover the contribution of the matrisome in orchestrating alterations in tissue ecology

391   during colitis, an integrative approach was used to map ECM ligands to receptors with altered

392   expression at the transcriptomic level. This ligand-receptor interaction analysis revealed

393   positive interactions between matrisomal proteins and T cell, B cell and macrophage

394   populations consistent with their expansion in inflamed tissues. Moreover, common

395   matrisomal protein subsets could be associated with specific cell types such as fibroblasts

396   and endothelial cells. For example, in the inflamed state, of the 38 proteins interacting with

397   fibroblasts and 39 interacting with endothelial cells, only Egfl7 is specific to endothelial cells.

398   The other 38 proteins interact with both endothelial cells and fibroblasts. Spatial analysis

399   showed a displacement of specific matrisomal ligands towards the top of crypts in inflamed

400   regions, which was associated with a displacement of both cell types, illustrating a co-

401   ordinated regulation driven by ECM remodelling accompanying colitis. Matrisomal ligands

402   associated with remodelling of these cell types were produced by immune cells or

403   inflammatory fibroblasts, representing an example of dynamic reciprocity in the context of

404   colitis.

405

406   In contrast, in the pre-pathological state, fewer matrisomal ligands and interactors are

407   engaged and they are predicted to have an impact on a smaller number of cell types, primarily

408   epithelial and endothelial cells. This suggests that assessing the pre-pathological tissue

409   landscape in terms of how altered ECM composition engages with and regulates different cell

410   types might provide new insights into how early tissue adaptations delay the subsequent onset

411   of colitis. These observations will benefit from performing complementary studies of

412   uninvolved mucosa in UC patients to establish if episodic recurrence of colitis in defined

413   regions relates to local loss of protective mechanisms rather than an acquisition of

414   susceptibility.

415

416   **Methods**

417   **Mice**

15

418   Animal care and procedures were performed in the Cancer Research UK Cambridge Institute

419   Biological Resource Unit according to UK Home Office guidelines. Mice are of C57BL/6

420   background. The *Muc2^KO* line used was described by Velcich *et al.*[25]. Genotyping of *Muc2^KO*

421   mice was outsourced to Transnetyx (Cordova, TN).

422   **Treatment of animals**. The mice were housed under controlled conditions (temperature (21

423   ± 2°C), humidity (55 ± 10%), 12h light/dark cycle) in a specific-pathogen-free (SPF) facility

424   (tested according to the recommendations for health monitoring by the Federation of

425   European Laboratory Animal Science Associations). Animals had unrestricted access to food

426   and water.

427

428   **Histology**

429   **Immunohistochemistry/immunofluorescence.** Mouse colons were opened longitudinally,

430   fixed overnight in 4% PFA and processed for histology by conventional means. Sections were

431   de-waxed and re-hydrated followed by heat-induced epitope-retrieval using 10 mM Tri-sodium

432   Citrate buffer pH6.0. Immunochemistry or immunofluorescence was performed as described

433   previously[48]. Antibodies used in immunohistochemistry were Muc2 (Santa Cruz, cat. sc-

434   15334) and Ki67 (Abcam, ab15580). Antibodies used in immunofluorescence were E-cadherin

435   (BD Bioscience cat. 610182), B220/CD45R (R&D systems, cat. MAB1217), Grem1 (R&D

436   systems, cat. AF956), CD68a (Cell signalling, cat. 98941), Serpinh1 (Abcam, cat. ab109117),

437   CD31 (Abcam, cat. ab28364) and Lamc1 (Abcam, cat. ab233389).

438   **Spatial phenotyping.** Fixed colons were swiss-rolled luminal side in, starting from the

439   proximal end. On-edge sections were cut at 3 different levels and stained with haematoxylin

440   and eosin. Immune hallmarks were scored and normalised along the colon length.

441   **Image analysis.** The Indica Labs HALO image analysis platform was used. The Indica Lab's

442   Area Quantification FL v.1.0 program was used for quantification of fluorescent staining

443   (percentage tissue stained and intensity of staining). The Indica Labs' Highplex FL v.2.2.3

444   program was used for cell deconvolution based on DAPI nuclear expression and quantification

445   of percentage of cells expressing specific fluorescent markers. For spatial analysis along the

446 crypt axis, layers of the same area were drawn at the bottom, middle and top of the crypt,

447 within which the analysis was performed.

448 **Extracellular matrix enrichment**

449 **Decellularization.** Colon tissue was dissected from 6-months-old *Muc2^hom^* mice and *Muc2^het^*

450 mice (n=5 mice per genotype). After flushing with PBS, 40-60mg colon tissue pellets were

451 flash-frozen and stored at -80C. On the day of decellularization, pellets were homogenised

452 using a Qiagen Lyser, before processing using a Compartmental Extraction Kit (Millipore,

453 #2145) as per manufacturer's protocol. This allows extraction of extracellular matrix proteins

454 through stepwise washes with salt solutions and detergents.

455 **Western Blot.** Purification of each cellular compartment was tested by Western Blot with

456 antibodies for Gapdh (Sigma, cat. G8795), Hmga1 (Abcam, cat. ab129153), β1 integrin (BD

457 bioscience, cat. 610467), βactin (Abcam, cat. ab6276) and collagen1 (Invitrogen, cat. PA5-

458 95137). 1/5th of the extracellular protein pellet was dissolved in 30ul of 4x SDS buffer

459 supplemented with 100mM DTT. For other compartments, 10ul of sample was mixed with 10x

460 LDS and 4x SDS.

461

462 **Label-free Mass Spectrometry, DIA-MS**

463 **Chemicals.** Chemicals LC-MS-grade acetonitrile (ACN), methanol and water were obtained

464 from Burdick & Jackson (Muskegon, MI). Reagents for protein chemistry, including sodium

465 dodecyl sulphate (SDS), ammonium bicarbonate, triethylammonium bicarbonate (TEAB),

466 iodoacetamide (IAA), dithiothreitol (DTT), sequencing-grade endoproteinase Lys-C, and

467 formic acid (FA) were purchased from Sigma-Aldrich (St. Louis, MO). Sequencing-grade

468 trypsin was purchased from Promega (Madison, WI). Glycerol-free PNGase F was purchased

469 from New England BioLabs (Ipswich, MA).

470 **Solubilization of ECM Proteins.** The extracted ECM pellets were solubilized by agitation for

471 10 minutes in a solution containing 1% SDS and 50 mM DTT, followed by sonication for 10

472 minutes, and finally heating at 85 ºC for 1 hour with agitation.

473   **Protein Digestion and Desalting.** Samples were solubilized with 4% SDS and 50 mM TEAB

474   at pH 8. Proteins were reduced with 20 mM DTT (10 min at 50°C followed by 10 min at RT)

475   and then alkylated with 40 mM iodoacetamide (30 min at RT in the dark). Samples were

476   acidified with a final concentration of 1.2% phosphoric acid, and diluted with seven volumes

477   S-trap buffer (90% methanol in 100 mM TEAB, pH 8). Samples were then loaded onto the S-

478   trap micro spin columns (Protifi, Farmingdale, NY) and spun at 4,000 x g for 10 seconds. The

479   S-Trap columns were washed with S-Trap buffer twice at 4,000 x g for 10 seconds each,

480   before incubating the proteins with 250 ng of Sequencing-Grade Endoproteinase Lys-C in 50

481   mM TEAB (pH 8) at 37°C for 2 hours. Then, 2.4µg of sequencing grade trypsin in 50 mM

482   TEAB (pH 8) for 1 hour at 47°C were added to the sample. After 1-hour digestion at 47°C, the

483   same amount of trypsin was added again and proteins were digested overnight at 37°C.

484   Peptides were sequentially eluted with 50 mM TEAB (pH 8), 0.5% FA in water, and 50% CAN,

485   0.5% FA in water. After vacuum drying, samples were resuspended in 300µL of 25 mM

486   ammonium bicarbonate in water, and spot-checked to ensure a pH of 7-8. Subsequently, 9µl

487   (4,500 U) of glycerol-free PNGase F were added, and samples were incubated for 3 hours at

488   37ºC with agitation. This reaction was quenched with 10% FA in water for a final concentration

489   of 1%, and spot-checked again to ensure a pH of 2–3. The quenched peptide samples were

490   vacuum dried and resuspended in 20µL of 0.2% FA, before desalting them using ziptips with

491   0.6µL C_{18} resin (Sigma-Aldrich). Finally, samples were concentrated in a vacuum concentrator

492   and re-suspended in aqueous 0.2% FA containing indexed retention time peptide standards

493   (iRT; Biognosys, Schlieren, Switzerland)[49].

494   **Mass Spectrometric Analysis.** LC-MS/MS analyses were performed on a Dionex UltiMate

495   3000 system coupled to an Orbitrap Eclipse Tribrid mass spectrometer (both from Thermo

496   Fisher Scientific, San Jose, CA). The solvent system consisted of 2% ACN, 0.1% FA in water

497   (solvent A) and 98% ACN, 0.1% FA in water (solvent B). Proteolytic peptides were loaded

498   onto an Acclaim PepMap 100 C_{18} trap column (0.1 x 20 mm, 5 µm particle size; Thermo Fisher

499   Scientific) over 5 min at 5 µL/min with 100% solvent A. Peptides were eluted on an Acclaim

500   PepMap 100 C_{18} analytical column (75 µm x 50 cm, 3 µm particle size; Thermo Fisher

501 Scientific) at 0.3 µL/min using the following gradient of solvent B: 2% for 5 min, linear from 2%

502 to 20% in 125 min, linear from 20% to 32% in 40 min, up to 80% in 1 min, 80% for 9 min, and

503 down to 2% in 1 min. The column was equilibrated at 2% for 29 min (total gradient length =

504 210 min).

505 Every sample was acquired in data-independent acquisition (DIA) mode [20-22] using the

506 following settings: full MS spectra were collected at 120,000 resolution (AGC target: 3e6 ions,

507 maximum injection time: 60ms, 350-1,650 m/z), and MS2 spectra at 30,000 resolution (AGC

508 target: 3e6 ions, maximum injection time: Auto, NCE: 27, fixed first mass 200 m/z). The

509 isolation scheme consisted in 26 variable windows covering the 350-1,650 m/z range with an

510 overlap of 1 m/z[22] (Supplemental Table S1).

511 **DIA-MS Data Processing with Spectronaut.** All DIA data was processed in Spectronaut

512 version 14.10.201222.47784 (Biognosys) using directDIA. Data was searched against the

513 *Mus musculus* proteome with 58,430 protein entries (UniProtKB-TrEMBL), accessed on

514 01/31/2018. Trypsin/P was set as digestion enzyme and two missed cleavages were allowed.

515 Cysteine carbamidomethylation was set as fixed modification, and methionine oxidation and

516 protein N-terminus acetylation as variable modifications. Data extraction parameters were

517 selected as dynamic, and non-linear iRT calibration with precision iRT was selected.

518 Identification was performed using a 1% precursor and protein q-value, and indexed retention

519 time (iRT) profiling was selected. Quantification was based on the MS/MS peak area of the 3-

520 6 best fragment ions per precursor ion, peptide abundances were obtained by summing

521 precursor abundances and protein abundances by summing peptide abundances.

522 Interference correction was selected, and local normalization was applied. Differential protein

523 abundance analysis was performed using paired t-test, and p-values were corrected for

524 multiple testing, specifically applying groupwise testing corrections using the Storey

525 method[50,51]. For differential analysis, protein groups with at least two unique peptides and q-

526 value ≤ 0.05 were considered to be significantly altered (Table S2).

527

528 **Bulk RNA sequencing**

529   Colon samples from three 10-months-old *Muc2^{hom}* and *Muc2^{het}* mice were used for bulk RNA

530   sequencing. RNA extraction was performed following instructions from the Quiagen RNA

531   extraction kit. Library prep was performed by *the* genomics core at the CRUK Cambridge

532   Institute using the Illumina Stranded mRNA Prep kit (Illumina, 20040532) according to the

533   manufacturer's instructions. Samples were submitted for sequencing in the Illumina Novaseq

534   platform with 50 bp paired end reads. Differential expression analysis was performed using

535   DESeq2. An interaction model was used to identify differentially expressed genes in proximal

536   or mid- *Muc2^{hom}* colon compared to location matched *Muc2^{het}* samples.

537

538

539   **Bioinformatic Analysis**

540   Plots were generated using either the ggplot2 package in R or Graphpad Prism.

541   **Pearson correlation.** The Pearson coefficients of correlation were determined between the

542   different replicates using the stats package in R (version 4.0.2; RStudio, version 1.3.1093)

543   and the abundances of all quantifiable protein groups as input.

544   **Pathway analysis.** An over-representation analysis was performed using

545   ConsensusPathDBhuman (Release 35, 05.06.2021)[52] to determine which gene ontology (GO)

546   terms were significantly enriched. GO terms identified from the over or under representation

547   analysis were subjected to the following filters: q-value < 0.01, number of background genes

548   ≥ 5, number of candidate genes ≥ 2, log2 fold enrichment >1 or <-1.

549   **Cell type deconvolution.** Cell type signatures were created by combining single-cell RNA

550   sequencing datasets provided in Kinchen *et al.,* 2018, Smillie *et al.,* 2019 and Mitsialis *et al.,*

551   2020 [DATASETS][39-41]. Two different datasets were created, with more or less specific cell

552   types (see supplemental Figure 3F and supplemental Figure 5B). Significantly differentially

553   expressed proteins (Qvalue < 0.05) or genes (from transcriptomic data, adjusted pvalue <

554   0.05) were mapped to cell type signatures and the spread of log2 fold changes for each cell

555   type was used to extrapolate information on enrichment.

556 **Protein interaction analysis**. The ligand-receptor interaction set from the

557 CellCellInteractions database (Version 1.0 - Built April 25, 2017, containing iRefIndex version

558 14, Pathway Commons version 8 and BioGRID version 3.4.147)[34] was used to find potential

559 interactors to matrisomal proteins.

560 **SLRPs pathway enrichment.** Pathway enrichment scores were calculated as:

561 $ES^{pathway} = ES^{interactors} * (N^{diff.exp.interactors}/N^{interactors})$

562 Where:

563   - $ES^{interactors}$ is calculated as average(Log2FC) * average(Qvalue) for all interators

564   found in the pathway

565   - $N^{diff.exp.interactors}$ represents the number of interactors to SLRPs mapped to the pathway

566   and differentially expressed in the relevant comparison

567   - $N^{interactors}$ represents the total number of potential interactors to all SLRPs with

568   interactors in the pathway

569 **Integrative network analysis.** Proteins part of the colon matrisome[53] and differentially

570 expressed in the comparison of either proximal inflamed vs true normal or middle matched

571 normal vs true normal were used for this analysis. Those datasets were processed to keep

572 proteins annotated as ligands (Bader Lab dataset of protein types). Bulk RNA sequencing

573 performed on $Muc2^{hom}$ vs $Muc2^{het}$ colon samples was used to identify differentially expressed

574 genes in the same comparisons. Those datasets were processed to keep genes annotated

575 as receptors, and mapped to matrisomal ligands with which they have a described interaction.

576 **Circos plots.** The circlize package was used to produce circos plots[54], where matrisomal

577 proteins were aligned on the bottom and linked to potential interactors at the top, which were

578 either mapped to pathways or cell types.

579

580 **References**

581   1 **Coussens, L. M. and Werb, Z**. (2018) 'Inflammation and cancer', *Environmental*

582   *Health and Preventive Medicine*, 23(1).

583   2   Pompili, S. *et al.* (2021) 'The Charming World of the Extracellular Matrix: A Dynamic
584   and Protective Network of the Intestinal Wall', *Frontiers in Medicine*, 8, pp. 1–19.

585   3   **McCabe, M. C., Saviola, A. J. and Hansen, K. C.** (2023) 'Mass Spectrometry-Based
586   Atlas of Extracellular Matrix Proteins across 25 Mouse Organs', *Journal of Proteome*
587   *Research*, 22(3), pp. 790–801.

588   4   **Sakakura, T., Nishizuka, Y. and Dawe, C. J.** (1976) 'Mesenchyme-dependent
589   morphogenesis and epithelium-specific cytodifferentiation in mouse mammary gland',
590   *Science*, 194(4272), pp. 1439–1441.

591   5   **Mott, J. D. and Werb, Z.** (2004) 'Regulation of matrix biology by matrix
592   metalloproteinases'. *Current Opinion in Cell Biology.* 16, pp. 558–564.

593   6   **Muncie, J. M. and Weaver, V. M.** (2019) 'The Physical and Biochemical Properties of
594   the Extracellular Matrix Regulate Cell Fate', *Curr Top Dev Biol.* 130, pp. 1–30.

595   7   Nakayama, K. H. *et al.* (2010) 'Decellularized Rhesus Monkey Kidney as a Three-
596   Dimensional Scaffold for Renal Tissue Engineering', *Tissue Engineering,* 16(7), pp.
597   2207-2216.

598   8   Rana, D. *et al.* (2017) 'Development of decellularized scaffolds for stem cell-driven
599   tissue engineering', *Journal of Tissue Engineering and Regenerative Medicine*, 11, pp.
600   942–965.

601   9   Ullah, I. *et al.* (2020) 'Adult Tissue Extracellular Matrix Determines Tissue Specification
602   of Human iPSC-Derived Embryonic Stage Mesodermal Precursor Cells'. *Advanced*
603   *Science,* 7, pp. 1-10.

604 10 **Bissell, M. J., Hall, H. G. and Parry, G.** (1982) 'How does the extracellular matrix
605   direct gene expression?', *Journal of Theoretical Biology*, 99(1), pp. 31–68.

606 11 Kirov, S. *et al.* (2019) 'Degradation of the extracellular matrix is part of the pathology
607   of ulcerative colitis', *Molecular Omics*, 15(1), pp. 67–76.

608 12 Baugh, M. D. *et al.* (1999) 'Matrix metalloproteinase levels are elevated in inflammatory
609   bowel disease', *Gastroenterology*, 117(4), pp. 814–822.

610   13 Ingber, D. E. (2002) 'Cancer as a disease of epithelial-mesenchymal interactions and
611   extracellular matrix regulation', *Differentiation*, 70(9–10), pp. 547–560.

612   14 **Pickup, M. W., Mouw, J. K. and Weaver, V. M.** (2014) 'The extracellular matrix
613   modulates the hallmarks of cancer'*, EMBO report*, 15(12), pp. 1243–1253.

614   15 Marangio A., *et al.* (2022). 'The Study of the Extracellular Matrix in Chronic
615   Inflammation: A Way to Prevent Cancer Initiation?' *Cancers,* 14, 5903.

616   16 Ordás, I. *et al.* (2012) 'Ulcerative colitis', *The Lancet*, 380(9853), pp. 1606–1619.

617   17 Mortensen J.H., *et al.* (2019). 'The intestinal tissue homeostasis –the role of
618   extracellular matrix remodeling in inflammatory bowel disease'. *Gastroenterology*
619   *Hepatology*. 10, pp. 977-993.

620   18 Koelink, P. J. *et al.* (2014) 'Collagen degradation and neutrophilic infiltration : a vicious
621   circle in inflammatory bowel disease', *Gut*. 63 pp. 578–587.

622   19 Gordon, I. O. *et al.* (2018) 'Fibrosis in ulcerative colitis is directly linked to severity and
623   chronicity of mucosal inflammation', *Alimentary Pharmacology and Therapeutics*,
624   47(7), pp. 922–939.

625   20 **Ullman, T. A. and Itzkowitz, S. H.** (2011) 'Intestinal Inflammation and Cancer',
626   *YGAST*, 140(6), pp. 1807-1816.

627   21 Gillet, L. C. *et al.* (2012) 'Targeted data extraction of the MS/MS spectra generated by
628   data-independent acquisition: A new concept for consistent and accurate proteome
629   analysis', *Molecular and Cellular Proteomics*, 11(6), pp. 1–17.

630   22 Collins, B. C. *et al.* (2017) 'Multi-laboratory assessment of reproducibility, qualitative
631   and quantitative performance of SWATH-mass spectrometry', *Nature*
632   *Communications*. Springer US, 8(1), pp. 1–11.

633   23 Bruderer R., et al (2017) 'Optimization of experimental parameters in data-independent
634   mass spectrometry significantly increases depth and reproducibility of results',
635   Molecular & Cellular Proteomics, 16(12), pp. 2296-309

636   24 Bons, J. *et al.* (2022) 'Data-independent acquisition and quantification of extracellular

637   matrix from human lung in chronic inflammation-associated carcinomas', *Proteomics*,

638   pp. 1-16.

639   25 Velcich, A. *et al.* (2002) 'Colorectal cancer in mice genetically deficient in the mucin

640   Muc2', *Science*, 295(5560), pp. 1726–1729.

641   26 Van der Sluis, M. *et al.* (2006) 'Muc2-Deficient Mice Spontaneously Develop Colitis,

642   Indicating That MUC2 Is Critical for Colonic Protection', *Gastroenterology*, 131(1), pp.

643   117–129.

644   27 Shao, X. *et al.* (2023) 'MatrisomeDB 2.0: 2023 updates to the ECM-protein knowledge

645   database', *Nucleic acids research*, 51, pp. D1519–D1530.

646   28 **Moll, R., Divo, M. and Langbein, L.** (2008) 'The human keratins: Biology and

647   pathology', *Histochemistry and Cell Biology*, 129(6), pp. 705–733.

648   29 **Hocking, A. M., Shinomura, T. and McQuillan, D. J.** (1998) 'Leucine-rich repeat

649   glycoproteins of the extracellular matrix', *Matrix Biology*, 17(1), pp. 1–19.

650   30 Czarnewski, P. *et al.* (2019) 'Conserved transcriptomic profile between mouse and

651   human colitis allows unsupervised patient stratification', *Nature Communications*,

652   10(1), pp. 1–11.

653   31 **Solé, M. and Unzeta, M.** (2011) 'Vascular cell lines expressing SSAO/VAP-1: a new

654   experimental tool to study its involvement in vascular diseases', *Biology of the Cell*,

655   103(11), pp. 543–557.

656   32 **Schaefer, L., Iozzo, R.** (2008) 'Biological functions of the small leucine-rich

657   proteoglycans: from genetics to signal transduction.' *The Journal of biological*

658   *chemistry* vol. 283,31, pp 21305-9.

659   33 **Iozzo, R. V. and Schaefer, L.** (2015) 'Proteoglycan form and function: A

660   comprehensive nomenclature of proteoglycans', *Matrix Biology*, 42, pp. 11–55.

661   34 **Isserlin, Ruth, Voisin, Veronique, Ailles, Laurie, & Bader, Gary D.** (2020). Cell-Cell

662   Interaction Database [DATASET]. Zenodo. https://doi.org/10.5281/zenodo.7589953

663   35   Shimshoni, E. *et al.* (2020) 'Distinct extracellular–matrix remodeling events precede
664   symptoms of inflammation', *Matrix Biology*, 96, pp. 47–68.

665   36   Johansson, M. E. V. *et al.* (2014) 'Bacteria penetrate the normally impenetrable inner
666   colon mucus layer in both murine colitis models and patients with ulcerative colitis',
667   *Gut*, 63(2), pp. 281–291.

668   37   Rifkin, D. B., Rifkin, W. J. and Zilberberg, L. (2018) 'LTBPs in biology and medicine:
669   LTBP diseases', *Matrix Biology*, 71–72, pp. 90–99.

670   38   Mak, J. W. Y. and Ng, S. C. (2020) 'Epidemiology of fibrostenosing inflammatory bowel
671   disease', *Journal of Digestive Diseases*, 21(6), pp. 332–335.

672   39   Kinchen, J. *et al.* (2018) 'Structural Remodeling of the Human Colonic Mesenchyme
673   in Inflammatory Bowel Disease', *Cell*, 175(2), pp. 372-386.e17.

674   40   Smillie, C. S. *et al.* (2019) 'Intra- and Inter-cellular Rewiring of the Human Colon during
675   Ulcerative Colitis', *Cell*, 178(3), pp. 714-730.

676   41   Mitsialis, V. *et al.* (2020) 'Single-Cell Analyses of Colon and Blood Reveal Distinct
677   Immune Cell Signatures of Ulcerative Colitis and Crohn's Disease', *Gastroenterology*,
678   159(2), pp. 591-608.

679   42   Troup S., *et al.* (2003) 'Reduced expression of the small leucine-rich proteoglycans,
680   lumican, and decorin is associated with poor outcome in node-negative invasive breast
681   cancer.' *Clin Cancer Res*, 9(1):207-14.

682   43   Charni, M. *et al.* (2017) 'P53 on the crossroad between regeneration and cancer', *Cell*
683   *Death and Differentiation*, 24(1), pp. 8–14.

684   44   **Fukushima, K. and Fiocchi, C.** (2004) 'Paradoxical decrease of mitochondrial DNA
685   deletions in epithelial cells of active ulcerative colitis patients', *American Journal of*
686   *Physiology - Gastrointestinal and Liver Physiology*, 286(5 49-5), pp. 804–813.

687   45   Beltran, B. *et al.* (2010) 'Mitochondrial dysfunction, persistent oxidative damage, and
688   catalase inhibition in immune cells of naïve and treated Crohn's disease', *Inflammatory*
689   *Bowel Diseases*, 16(1), pp. 76–86.

690   46 **Rath, E. and Haller, D**. (2012) 'Mitochondria at the Interface Between Danger

691   Signaling and Metabolism : Role of Unfolded Protein Responses in Chronic

692   inflammation', *Basic Science Review*, 18(7), pp. 1364–1377.

693   **47** Khaloian, S. *et al.* (2020) 'Mitochondrial impairment drives intestinal stem cell transition

694   into dysfunctional Paneth cells predicting Crohn's disease recurrence', *Gut*, 69(11),

695   pp. 1939–1951.

696   48 Tomic, G. *et al.* (2018) 'Phospho-regulation of ATOH1 Is Required for Plasticity of

697   Secretory Progenitors and Tissue Regeneration', *Cell Stem Cell*, 23(3), pp. 436--443.

698   49 Escher, C. *et al.* (2012). Using iRT, a normalized retention time for more targeted

699   measurement of peptides. *Proteomics*, *12*(8), pp. 1111–1121

700   50 Storey, J. D. (2002) 'A direct approach to false discovery rates'. *Journal of the Royal*

701   *Statistical Society: Series B (Statistical Methodology), 64*(3), pp. 479-498.

702   51 Burger, T. (2018). Gentle Introduction to the Statistical Foundations of False Discovery

703   Rate in Quantitative Proteomics. *J Proteome Res, 17*(1), pp. 12-22.

704   52 Kamburov, A. *et al.* (2009) 'ConsensusPathDB - A database for integrating human

705   functional interaction networks', *Nucleic Acids Research*, 37(SUPPL. 1), pp. 623–628.

706   53 Naba, A et al. (2012) 'The matrisome: in silico definition and in vivo characterization

707   by proteomics of normal and tumor extracellular matrices.' *Molecular & cellular*

708   *proteomics: MCP,* vol. 11,4, M111.014647.

709   54 Gu, Z. *et al.* (2014) 'Circlize implements and enhances circular visualization in R',

710   *Bioinformatics*, 30(19), pp. 2811–2812.

711   **Author names in bold designate shared co-first authorship**

712

713   **Figure legends**

714   **Figure 1: Proteomic analysis of an extracellular matrix-enriched tissue fraction reflects**

715   **disease state**

716   (A) Venn diagram showing the proportion of proteins identified through DIA-MS classed as

717   extracellular based on GO terms (953, 40.9% of the total protein fraction) and the proportion

26

718   of matrisomal proteins identified (162, amounting to 5% of the total protein fraction). (B)

719   Volcano plot displaying differentially expressed proteins in the proximal colon of *Muc2^hom* (n=4)

720   versus *Muc2^het* control (n=5) mice. Red indicates significant upregulation and blue significant

721   downregulation (Log2 FC >1 or <-1, Qvalue < 0.05). (C) Dot plot showing ConsensusPathDB

722   output of assessing pathway enrichment for *Muc2^hom* vs *Muc2^het* proximal colon samples. Top

723   significantly enriched pathways are represented (p value < 0.01). (D) Immunofluorescence

724   pictures showing enrichment for CD4+ (left) T cells and B220/CD45R+ (right) B cells in

725   *Muc2^hom* compared to Muc2^het proximal colon (DAPI marking nuclei and E-cadherin marking

726   colonic epithelial cells). (E) Violin plots showing the distribution of protein log2 fold changes in

727   *Muc2^hom* compared to Muc2^het proximal samples (Qvalue < 0.05), where proteins are mapped to

728   different cell type signatures (details in methods). Positive enrichment is indicated by the

729   median (black dot) sitting above the red 0 line, and negative enrichment when under this line.

730   (F) Protein differential expression signature extracted from the comparison of *Muc2^hom*

731   compared to Muc2^het proximal samples (see Supplemental Table S2). A subset is shown here

732   where log2 FC >2 or <-2 and Qvalue < 5x10e-7. (G) Box plots displaying differential

733   expression of proteins in *Muc2^hom* compared to Muc2^het proximal samples (Qvalue < 0.05),

734   where proteins are classed into extracellular matrix protein categories (*pvalue < 0.05,

735   pairwise t-test).

736

737   **Figure 2: Mid-colon regions in mature *Muc2^KO* mice are in a pre-pathological state**

738   (A) Venn diagram showing overlap in proteins up- and down regulated (Qvalue < 0.05) in

739   proximal and middle mature *Muc2^hom* mouse colon (inflamed and histologically normal

740   respectively, n=4 samples per condition) compared to location-matched *Muc2^het* control

741   samples (n=5 for each location). Circled in grey are proteins sharing the same direction of

742   change in both comparisons (see Supplemental Table S2, shared signature), circled in black

743   are proteins showing opposite direction of change (see Supplemental Table S2, opposite

744   signature). Fill colour from blue to red: lower to higher number of proteins in each subset of

745   the Venn diagram. (B) Bar plot displaying a subset of proteins extracted from the "shared

27

746   signature" (Qvalue < 0.01, log2 fold change >1 or <-1). Highlighted proteins are shown in C

747   and D. (C) Bar plots showing normalised protein abundance for Sulfotransferase Family 1A

748   Member 1 (Sult1a1), downregulated in both proximal and middle mature *Muc2^hom* compared

749   to location matched *Muc2^het* control samples. (D) Bar plots representing normalised protein

750   abundance for Gremlin (Grem1), showing a similar pattern to Sult1a1 with a stronger

751   sequential upregulation from mid- to proximal *Muc2^hom* mouse colon. (E) Representative

752   immunofluorescence images of Grem1 (in red) in *Muc2^het*, mid-colon *Muc2^hom* and proximal

753   *Muc2^hom* tissue. DAPI marks nuclei, E-cadherin marks colonic epithelial cells. The bottom

754   panel is a magnification of the white boxes in the top panel. (F) Bar plot displaying a subset

755   of proteins showing "opposing signatures" (Qvalue < 0.05, log2 fold change >1 or <-1).

756   Highlighted proteins are shown in G and H. (G) Bar plots showing normalised protein

757   abundance for Chemokine (C-C motif) ligand 6 (Ccl6), down-regulated in *Muc2^hom* mid-colon

758   and up-regulated in *Muc2^hom* proximal colon compared to *Muc2^het* control samples. (H)

759   Normalised protein abundance for Proline and arginine rich end leucine rich repeat protein

760   (Prelp), up-regulated in Muc2^hom mid-colon and down-regulated in *Muc2^hom* proximal colon

761   compared to *Muc2^het* control samples.

762

763   **Figure 3: Differential expression of proteoglycans distinguishes tissues in pre- and**

764   **pathological states**

765   (A) Bar plot showing distribution of log2 fold changes for Small-Leucine-Rich Proteoglycans

766   (SLRPs) found significantly differentially expressed in the comparison of middle *Muc2^hom* vs

767   *Muc2^het* and/or proximal *Muc2^hom*   vs *Muc2^het* samples (Qvalue < 0.05). (B) Box plots

768   representing differential protein expression as in (A), showing a downregulation from the

769   comparison of middle *Muc2^hom* vs *Muc2^het* to proximal *Muc2^hom* vs *Muc2^het* for SLRPs identified

770   in both comparisons. Dcn and Prelp are highlighted as candidates upregulated in the middle

771   *Muc2^hom* vs *Muc2^het* and downregulated in proximal *Muc2^hom* vs *Muc2^het*. (C) Bar plots showing

772   number of known interactors to SLRPs identified (left, in black) and those differentially

773   expressed in the comparison of proximal *Muc2^hom* vs *Muc2^het* or middle *Muc2^hom* vs *Muc2^het*

774 samples (right, in grey). (D) Circos plots displaying SLRPs on the bottom, differentially

775 expressed in proximal *Muc2^{hom}* vs *Muc2^{het}* samples (left) or middle *Muc2^{hom}* vs *Muc2^{het}*

776 samples (right). SLRPs are mapped to hallmark pathways (top) involving their known

777 interactors, enriched (see methods) either inproximal *Muc2^{hom}* vs *Muc2^{het}* samples (left) or

778 middle *Muc2^{hom}* vs *Muc2^{het}* samples (right).

779

780 **Figure 4: Matrisomal orchestration of tissue remodelling in colitis**

781 (A) Circos plot showing matrisomal ligands differentially expressed (Qvalue < 0.05) in proximal

782 *Muc2^{hom}* vs *Muc2^{het}* (left) or middle *Muc2^{hom}* vs *Muc2^{het}* (right), organised by groups (bottom)

783 linking cell types (top) containing receptors with which they are known to interact and are

784 differentially expressed at the transcriptomic level (adjusted p value < 0.05, see method for

785 details). Mean log2 fold change for interactors to matrisomal ligands mapping to each cell type

786 is shown as heat on the top panel. (B) Heatmap displaying matrisomal proteins part of the

787 inflamed matrisome (y axis) and the cell types within which they have potential interactors (x-

788 axis). The heat represents the number of interactors of each matrisomal ligand within each

789 cell type, and is normalised per row giving more power to unique interactions. (C)

790 Immunofluorescence images showing the location of CD31+ endothelial cells (left panel, in

791 red), CD68+ macrophages (middle panel, in yellow) and Serpinh1+ fibroblasts (right panel, in

792 green) in proximal and middle *Muc2^{hom}* colon. (D) Heatmaps showing the percentage of

793 positive cells for each of the 3 cell types in 3 regions of the same area along the crypt axis

794 (bottom, middle and top) in proximal and middle *Muc2^{hom}* colon (n=3). (E) Heatmaps showing

795 the expression of 4 ECM ligands in proximal and middle *Muc2^{hom}* colon, expressed as intensity

796 per area (n=3). (D) Representative immunofluorescence image showing the distribution of

797 Lamc1 along the crypt axis in proximal and middle *Muc2^{hom}* colon.

798

799 **Supplemental figure legends**

800 **Supplemental Figure 1: Characterisation of the *Muc2^{KO}* mouse model of colitis reveals**

801 **spatial differences in pathology**

802   (A) Immunohistochemistry images for Muc2 (top), Alcian Blue/PAS (middle) and Ki67

803   (bottom), showing Muc2 deletion, hindered mucin production and increased Ki67 expression

804   in *Muc2hom* compared to *Muc2wt* and *Muc2het* mouse colon. (B) Bar plots showing average crypt

805   length (left), as well as number of immune infiltrates (middle) and abscesses (right) per mm is

806   increased in *Muc2hom* compared to *Muc2het* colon (n=5 *Muc2hom* mice, median survival = 5.3

807   months, n=3 *Muc2het* mice, median survival= 4.5 months. Statistical significance was

808   determined using a Mann-Whitney test, * = p value<0.04). (C) Bar plots showing the presence

809   of pre-neoplastic hallmarks is restricted to *Muc2hom* colon. Dysplasia (left) and invasive length

810   (right) are quantified as proportion of respective pathology over the colon length, with E-

811   cadherin+ glands crossing the *muscularis mucosae* considered as invasive. (D)

812   Representative H&E images showing the spatial aspect of pathology along the mature

813   *Muc2hom* mouse colon. Proximal and distal regions are commonly inflamed, while the middle

814   colon remains histologically normal. (E) Mature *Muc2hom* colon drawing and histograms

815   representing the spatial heterogeneity seen in crypt length, number of immune infiltrates and

816   abscesses along the length of the colon (n=5 *Muc2hom* mice used, normalisation to n=3 *Muc2het*

817   mice. Quantification done over 12 colonic regions as shown on the y axis, error bars = 95%

818   confidence interval).

819

820   **Supplemental Figure 2: Timeline and sample location for ECM proteomics**

821   (A) Timeline sowing age of mice used for histopathological analysis and ECM proteomics. (B)

822   Scheme representing the location of samples used for ECM proteomics and associated level

823   of pathology.

824

825   **Supplemental Figure 3: Experimental approach to investigate ECM remodelling in**

826   **colitis**

827   (A) Experimental outline for label-free mass-spectrometry on decellularized fractions of

828   *Muc2KO* mouse colon. (B) Western blot showing sequential purification of the extracellular

829   tissue fraction. Gapdh was used to probe for removal of cytoplasmic proteins, Hmga1 of

830   nuclear proteins, β1-integrin of membrane proteins and Actin of cytoskeletal proteins.

831   Enrichment for ECM is seen with appearance of a band for Collagen1. (C) Heatmap displaying

832   identified and quantified protein groups for all samples, showing 2 out-layers outlined in black

833   that were removed from the analysis. (D) Box plots representing the distribution of Pearson

834   correlation coefficients for samples in each condition. (E) Principal Component Analysis using

835   the abundance of all quantified proteins between sample groups, where PC1 (x-axis)

836   separates samples based on genotype and PC2 (y-axis) based on location along the colon.

837   (F) Stacked bar plots showing the number of genes present in each extended cell type

838   signatures, coming from each of 3 colon single-cell RNA sequencing published datasets (see

839   details in methods).

840

841   **Supplemental Figure 4: *Muc2^KO* mice evolve from discrete to pan-colitis with age**

842   (A) Dotplot displaying log2 fold changes for each protein differentially expressed in *Muc2^hom*

843   mid- colon (y-axis) and proximal (x-axis) samples compared to location matched *Muc2^het* colon

844   sample (Qvalue < 0.05). (B) Survival analysis showing heterogeneity in *Muc2^hom* mice

845   survival, from 100 days up to 1-year-old (study endpoint). (C) Bar plot showing crypt height

846   for young (in grey, median age 1.8 month) and old (in pink, median age 12.4 months) *Muc2^hom*

847   mice, normalised to age matched *Muc2^het* controls (n=3 mice for each condition). (D)

848   Representative H&E pictures showing development of pathology in the proximal colon in

849   young *Muc2^hom* mice, which extends to the whole colon length in old *Muc2^hom* mice.

850

851   **Supplemental Figure 5: Building networks representing matrisomal orchestration of**

852   **cell type remodelling**

853   (A) Circos plot showing differentially expressed matrisomal ligands (bottom, Qvalue < 0.05)

854   and their known interactors differentially expressed at the transcriptomic level (top, adjusted

855   p value < 0.05), in proximal Muc2^hom (left) and middle *Muc2^hom* (right) compared to location

856   matched *Muc2^het* control samples. (B) Histograms showing the number of receptors

857   differentially expressed at the transcriptomic level (x a-axis, adjusted p value < 0.05), for each

858   matrisomal ligand differentially expressed in the comparison of proximal Muc2^{hom} vs $Muc2^{het}$

859   (left) or middle $Muc2^{hom}$ vs $Muc2^{het}$ (right). (C) Bar plot representing the number of genes

860   present in each combined cell type signatures, where discrete cell types were combined under

861   a general cell type (see method). (D) Histogram showing the number (x-axis) and name (figure

862   legend) of matrisomal proteins (as described in B) controlling each cell type present on the y-

863   axis (combined cell type signature). (E) Histogram displaying the number of cell types

864   (combined cell type signature) controlled by each matrisomal protein as described in B.

865

866   **Supplementary material**

867   **Tables**

868   Excel table S1: Tabel Sx_DIAIsolationScheme

869   Excel table S2: Candidates_2021_0623_EM2_directDIA_v4_MinusEM2_14-

870   17_Processed_2023_-510_v3

871   Excel table S3: Updated colon matrisome

872   Excel table S4: Signatures
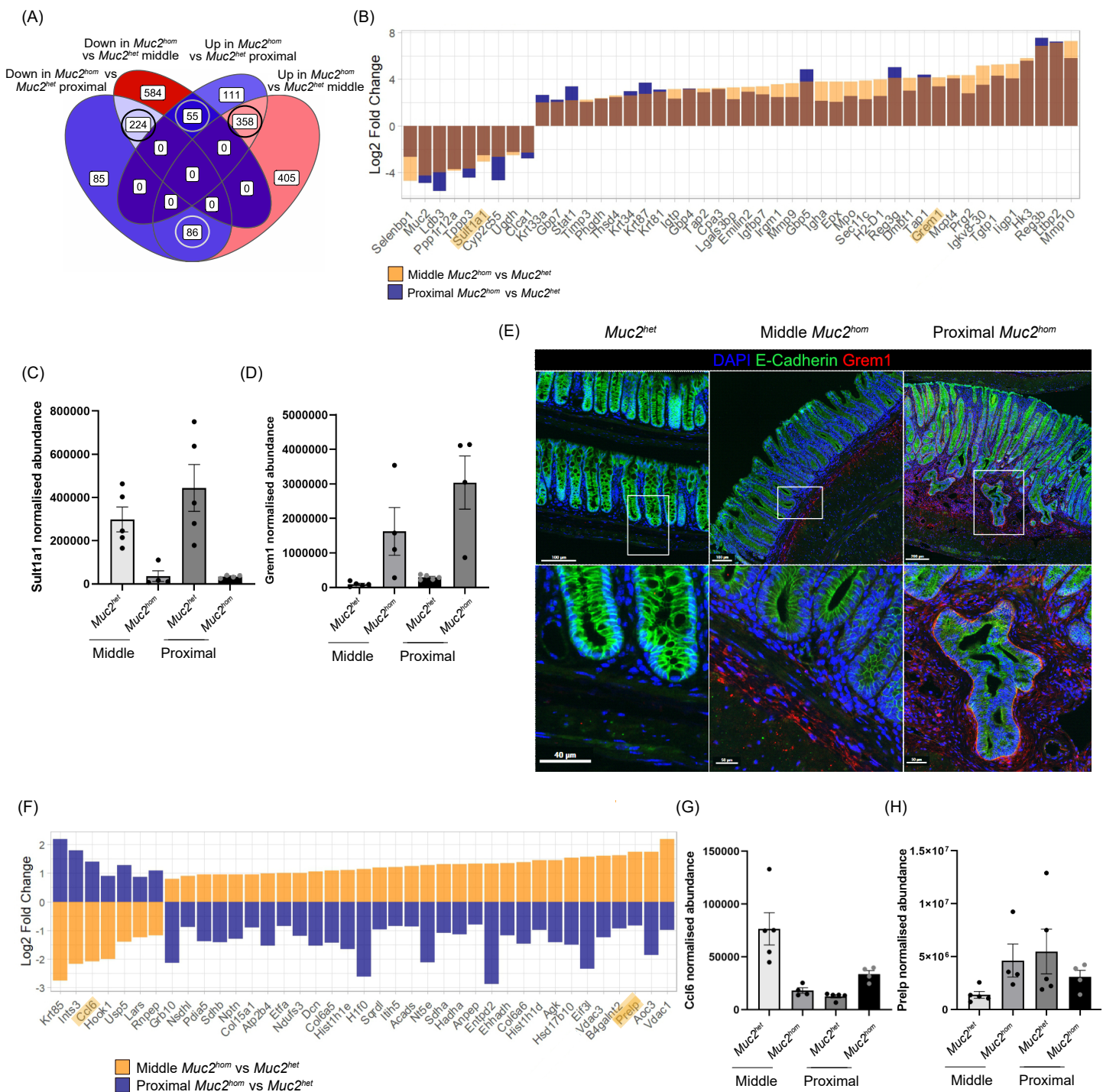
873   Excel table S5: Networks

Figure 1: Proteomic analysis of an extracellular matrix-enriched tissue fraction reveals disease state
(A) Venn diagram showing the proportion of proteins identified through DIA-MS classed as extracellular based on GO terms (953, 40.9% of the total protein fraction) and the proportion of matrisomal proteins identified (162, amounting to 5% of the total protein fraction). (B) Volcano plot displaying differentially expressed proteins in the proximal colon of *Muc2hom* (n=4) versus *Muc2het* control (n=5) mice. Red indicates significant upregulation and blue significant downregulation (Log2 FC >1 or <-1, Qvalue < 0.05). (C) Dot plot showing ConsensusPathDB output of assessing pathway enrichment for *Muc2hom* vs *Muc2het* proximal colon samples. Top significantly enriched pathways are represented (p value < 0.01). (D) Immunofluorescence pictures showing enrichment for CD4+ (left) T cells and B220/CD45R+ (right) B cells in *Muc2hom* compared to *Muc2het* proximal colon (DAPI marking nuclei and E-cadherin marking colonic epithelial cells). (E) Violin plots showing the distribution of protein log2 fold changes in *Muc2hom* compared to *Muc2het* proximal samples (Qvalue < 0.05), where proteins are mapped to different cell type signatures (details in methods). Positive enrichment is indicated by the median (black dot) sitting above the red 0 line, and negative enrichment when under this line. (F) Protein differential expression signature extracted from the comparison of *Muc2hom* compared to *Muc2het* proximal samples (see Supplemental Table S2). A subset is shown here where log2 FC >2 or <-2 and Qvalue < 5x10e-7. (G) Box plots displaying differential expression of proteins in *Muc2hom* compared to *Muc2het* proximal samples (Qvalue < 0.05), where proteins are classed into extracellular matrix protein categories (*pvalue < 0.05, pairwise t-test).

Figure 2: Mid-colon regions in mature *Muc2^KO* mice are in a pre-pathological state
(A) Venn diagram showing overlap in proteins up- and down regulated (Qvalue < 0.05) in proximal and middle mature *Muc2^hom* mouse colon (inflamed and histologically normal respectively, n=4 samples per condition) compared to location-matched *Muc2^het* control samples (n=5 for each location). Circled in grey are proteins sharing the same direction of change in both comparisons (see Supplemental Table S2, shared signature), circled in black are proteins showing opposite direction of change (see Supplemental Table S2, opposite signature). Fill colour from blue to red: lower to higher number of proteins in each subset of the Venn diagram. (B) Bar plot displaying a subset of proteins extracted from the "shared signature" (Qvalue < 0.01, log2 fold change >1 or <-1). Highlighted proteins are shown in C and D. (C) Bar plots showing normalised protein abundance for Sulfotransferase Family 1A Member 1 (Sult1a1), downregulated in both proximal and middle mature *Muc2^hom* compared to location-matched *Muc2^het* control samples. (D) Bar plots representing normalised protein abundance for Gremlin (Grem1), showing a similar pattern to Sult1a1 with a stronger sequential upregulation from mid- to proximal *Muc2^hom* mouse colon. (E) Representative immunofluorescence images of Grem1 (in red) in *Muc2^het*, mid-colon *Muc2^hom* and proximal *Muc2^hom* tissue. DAPI marks nuclei, E-cadherin marks colonic epithelial cells. The bottom panel is a magnification of the white boxes in the top panel. (F) Bar plot displaying a subset of proteins showing "opposing signatures" (Qvalue < 0.05, log2 fold change >1 or <-1). Highlighted proteins are shown in G and H. (G) Bar plots showing normalised protein abundance for Chemokine (C-C motif) ligand 6 (Ccl6), down-regulated in *Muc2^hom* mid-colon and up-regulated in *Muc2^hom* proximal colon compared to *Muc2^het* control samples. (H) Normalised protein abundance for Proline and arginine rich end leucine rich repeat protein (Prelp), up-regulated in *Muc2^hom* mid-colon and down-regulated in *Muc2^hom* proximal colon compared to *Muc2^het* control samples.
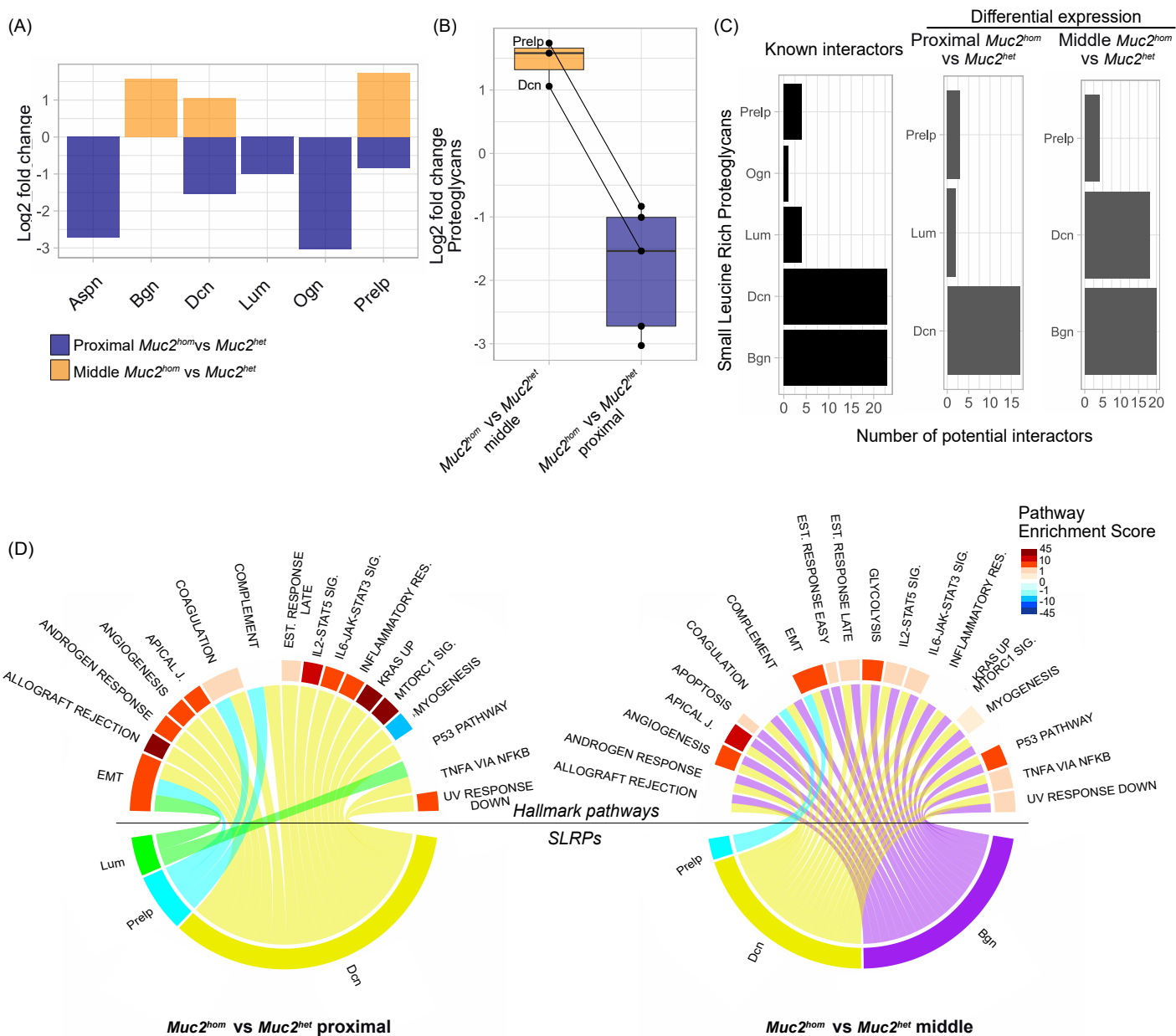
Figure 3: Differential expression of proteoglycans distinguishes tissues in pre- and pathological states

(A) Bar plot showing distribution of log2 fold changes for Small-Leucine-Rich Proteoglycans (SLRPs) found significantly differentially expressed in the comparison of middle $Muc2^{hom}$ vs $Muc2^{het}$ and/or proximal $Muc2^{hom}$ vs $Muc2^{het}$ samples (Qvalue < 0.05). (B) Box plots representing differential protein expression as in (A), showing a downregulation from the comparison of middle $Muc2^{hom}$ vs $Muc2^{het}$ to proximal $Muc2^{hom}$ vs $Muc2^{het}$ for SLRPs identified in both comparisons. Dcn and Prelp are highlighted as candidates upregulated in the middle $Muc2^{hom}$ vs $Muc2^{het}$ and downregulated in proximal $Muc2^{hom}$ vs $Muc2^{het}$. (C) Bar plots showing number of known interactors to SLRPs identified (left, in black) and those differentially expressed in the comparison of proximal $Muc2^{hom}$ vs $Muc2^{het}$ or middle $Muc2^{hom}$ vs $Muc2^{het}$ samples (right, in grey). (D) Circos plots displaying SLRPs on the bottom, differentially expressed in proximal $Muc2^{hom}$ vs $Muc2^{het}$ samples (left) or middle $Muc2^{hom}$ vs $Muc2^{het}$ samples (right). SLRPs are mapped to hallmark pathways (top) involving their known interactors, enriched (see methods) either inproximal $Muc2^{hom}$ vs $Muc2^{het}$ samples (left) or middle $Muc2^{hom}$ vs $Muc2^{het}$ samples (right).
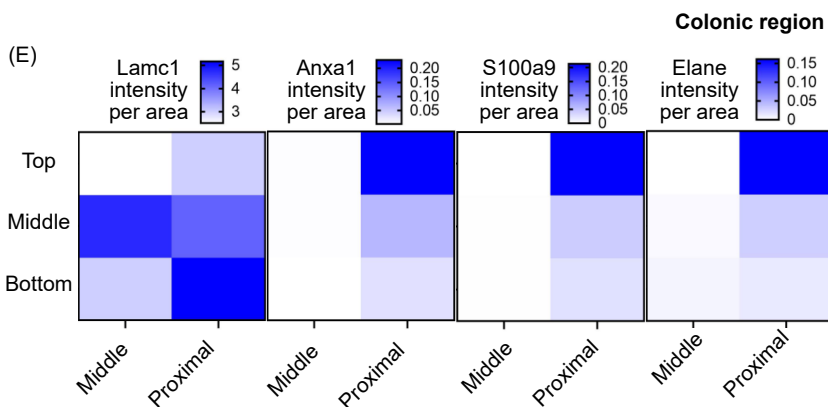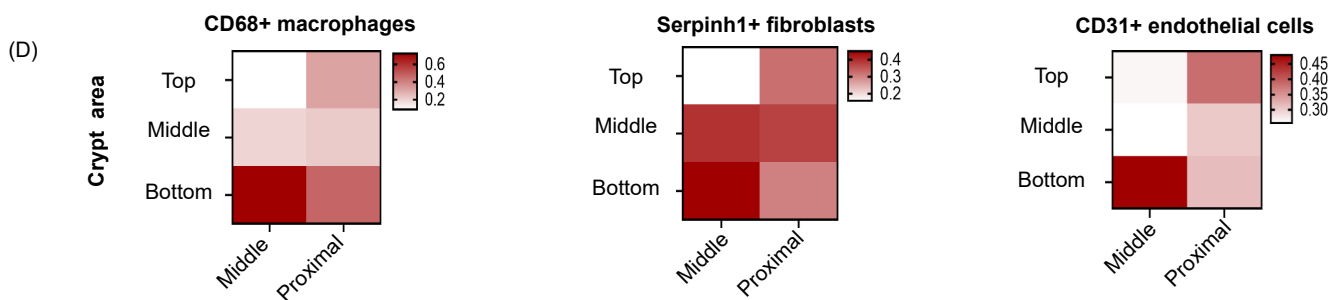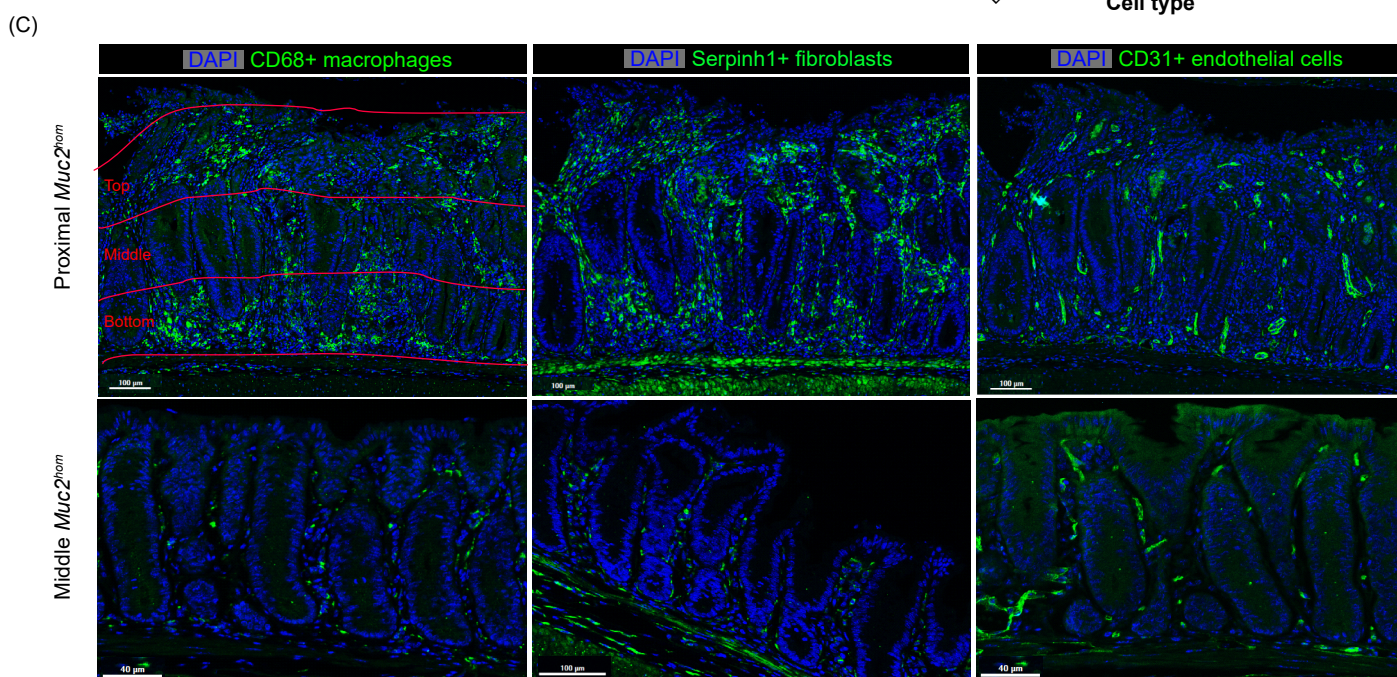
(A) *Muc2^{hom}* vs *Muc2^{het}* proximal   *Muc2^{hom}* vs *Muc2^{het}* middle

Key
- ECM regulators
- ECM associated
- Secreted factors
- Proteoglycans
- ECM glycoproteins
- Collagens

Mean log2 FC potential interactors
2.5
1
0

(B) Number of interactors and specificity of interaction
Low — High

(C)

(D) CD68+ macrophages   Serpinh1+ fibroblasts   CD31+ endothelial cells

(E) Lamc1 intensity per area   Anxa1 intensity per area   S100a9 intensity per area   Elane intensity per area

(F) Middle *Muc2^{hom}*   Proximal *Muc2^{hom}*

(Figure legend on next page)
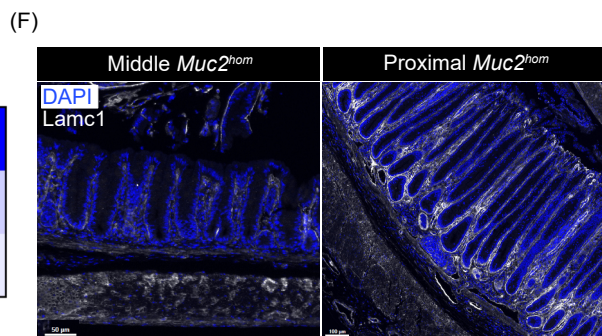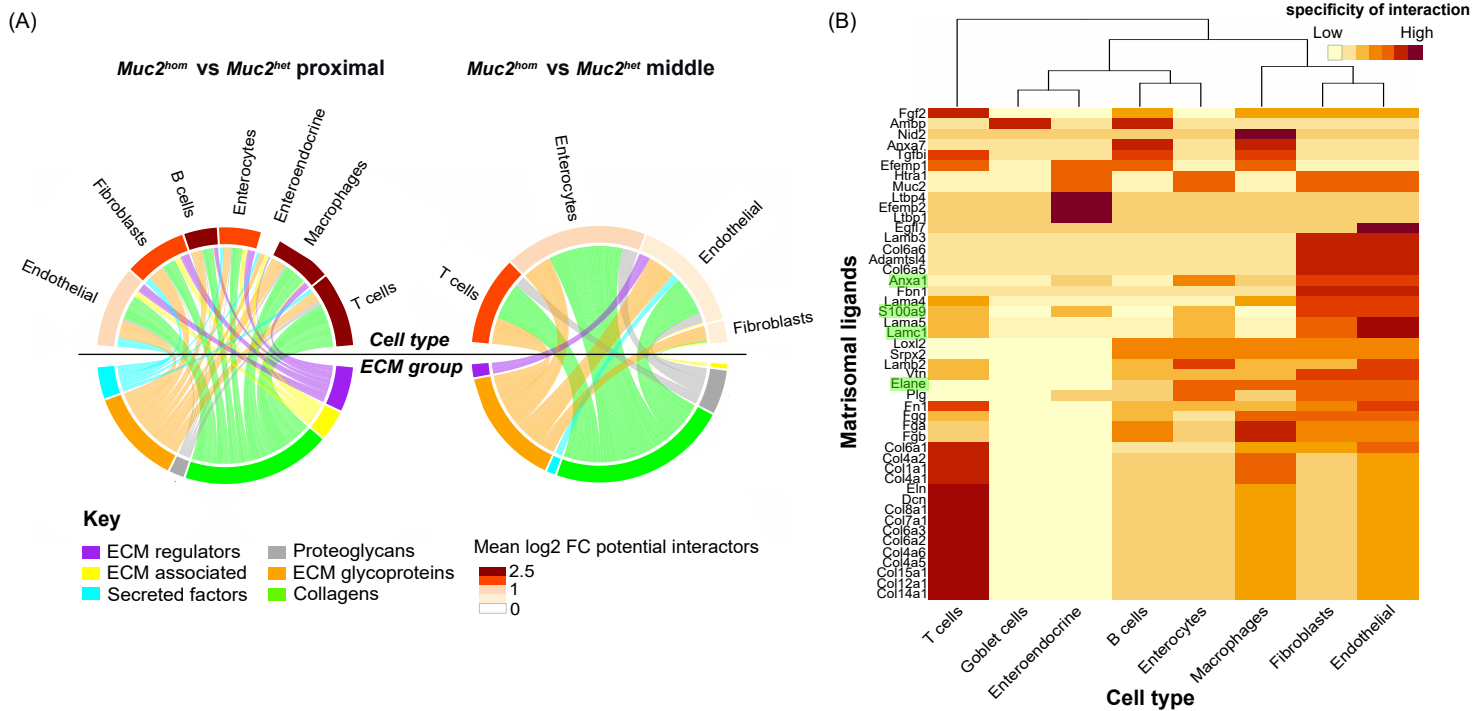
Figure 4: Matrisomal orchestration of tissue remodelling in colitis

(A) Circos plot showing matrisomal ligands differentially expressed (Qvalue < 0.05) in proximal *Muc2*^{hom} vs *Muc2*^{het} (left) or middle *Muc2*^{hom} vs *Muc2*^{het} (right), organised by groups (bottom) linking cell types (top) containing receptors with which they are known to interact and are differentially expressed at the transcriptomic level (adjusted p value < 0.05, see method for details). Mean log2 fold change for interactors to matrisomal ligands mapping to each cell type is shown as heat on the top panel. (B) Heatmap displaying matrisomal proteins part of the inflamed matrisome (y axis) and the cell types within which they have potential interactors (x-axis). The heat represents the number of interactors of each matrisomal ligand within each cell type, and is normalised per row giving more power to unique interactions. (C) Immunofluorescence images showing the location of CD31+ endothelial cells (left panel, in red), CD68+ macrophages (middle panel, in yellow) and Serpinh1+ fibroblasts (right panel, in green) in proximal and middle *Muc2*^{hom} colon. (D) Heatmaps showing the percentage of positive cells for each of the 3 cell types in 3 regions of the same area along the crypt axis (bottom, middle and top) in proximal and middle *Muc2*^{hom} colon (n=3). (E) Heatmaps showing the expression of 4 ECM ligands in proximal and middle *Muc2*^{hom} colon, expressed as intensity per area (n=3). (D) Representative immunofluorescence image showing the distribution of Lamc1 along the crypt axis in proximal and middle *Muc2*^{hom} colon.
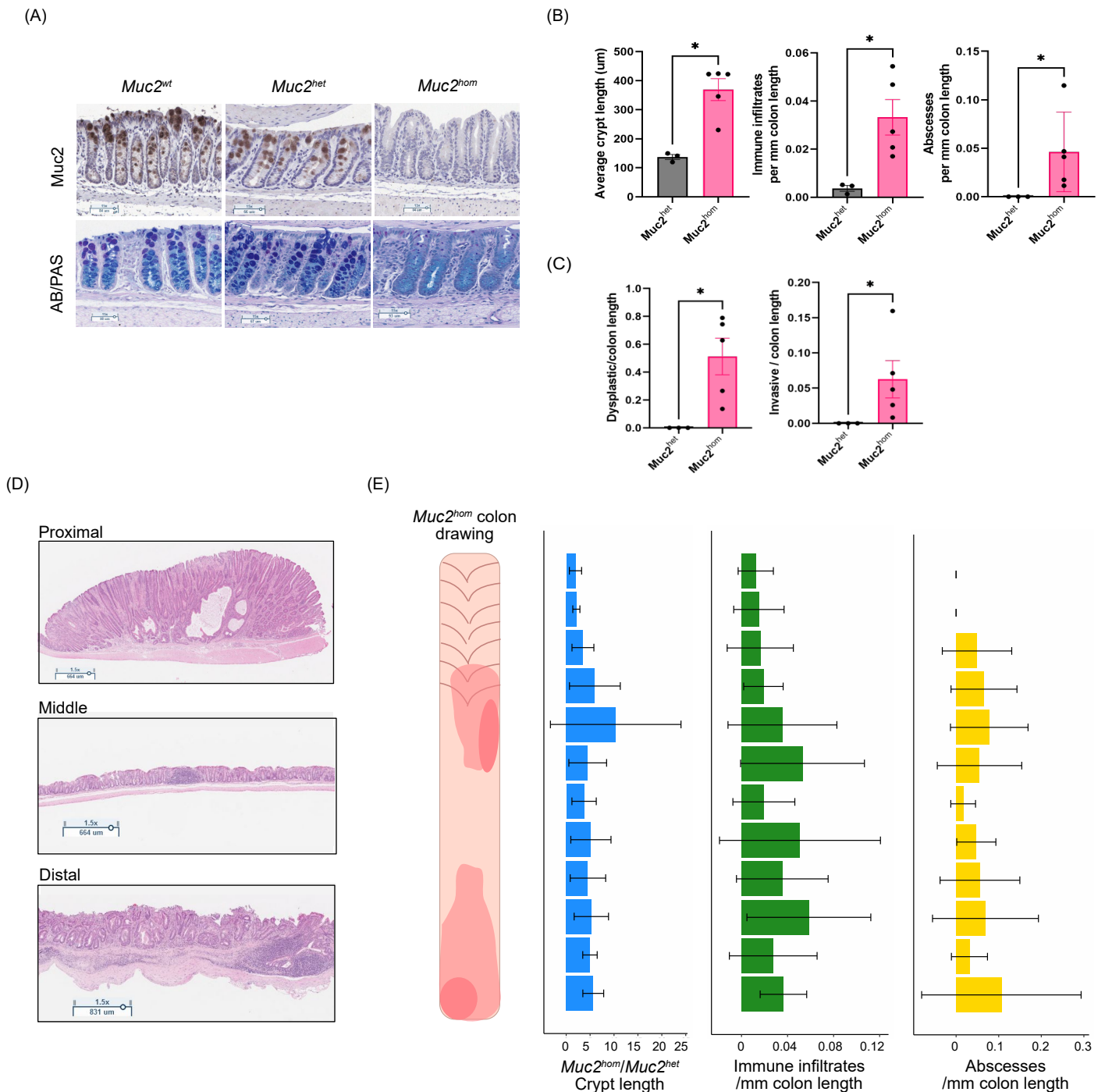
Supplemental Figure 1: Characterisation of the *Muc2^KO* mouse model of colitis reveals spatial differences in pathology

(A) Immunohistochemistry images for Muc2 (top), Alcian Blue/PAS (bottom), showing Muc2 deletion, hindered mucin production in *Muc2^hom* compared to *Muc2^wt* and *Muc2^het* mouse colon. (B) Bar plots showing average crypt length (left), as well as number of immune infiltrates (middle) and abscesses (right) per mm is increased in *Muc2^hom* compared to *Muc2^het* colon (n=5 *Muc2^hom* mice, median survival = 5.3 months, n=3 *Muc2^het* mice, median survival= 4.5 months. Statistical significance was determined using a Mann-Whitney test, * = p value<0.04). (C) Bar plots showing the presence of pre-neoplastic hallmarks is restricted to *Muc2^hom* colon. Dysplasia (left) and invasive length (right) are quantified as proportion of respective pathology over the colon length, with E-cadherin+ glands crossing the *muscularis mucosae* considered as invasive. (D) Representative H&E images showing the spatial aspect of pathology along the mature *Muc2^hom* mouse colon. Proximal and distal regions are commonly inflamed, while the middle colon remains histologically normal. (E) Mature *Muc2^hom* colon drawing and histograms representing the spatial heterogeneity seen in crypt length, number of immune infiltrates and abscesses along the length of the colon (n=5 *Muc2^hom* mice used, normalisation to n=3 *Muc2^het* mice. Quantification done over 12 colonic regions as shown on the y axis, error bars = 95% confidence interval).
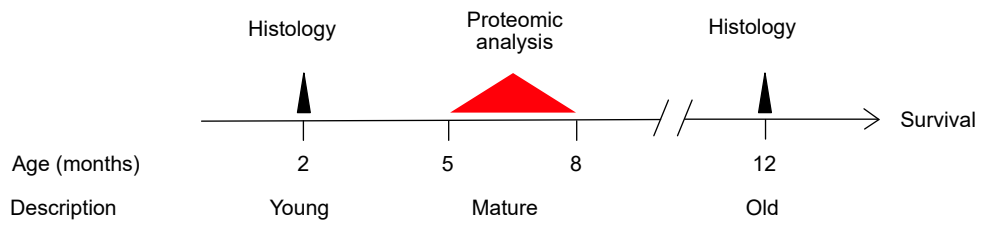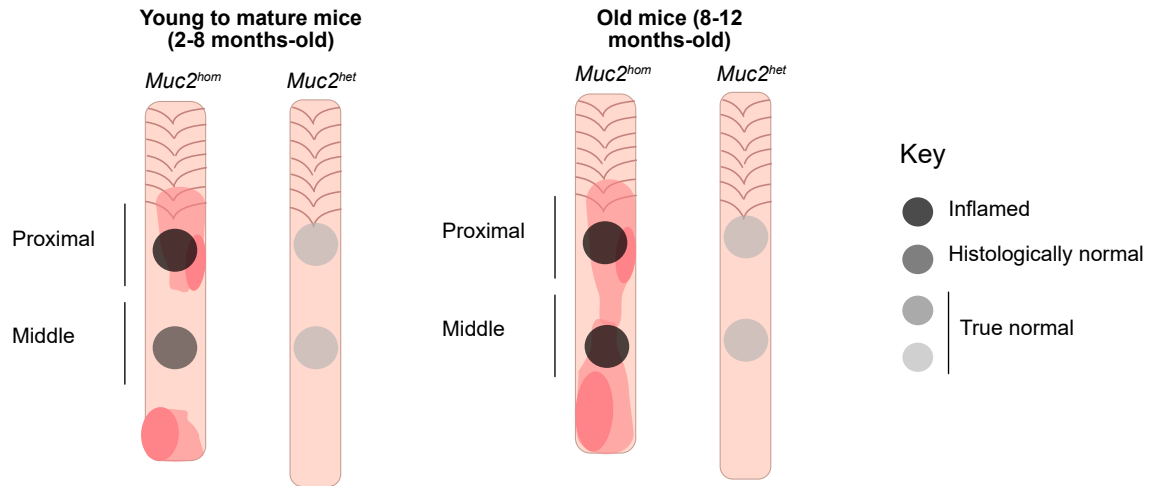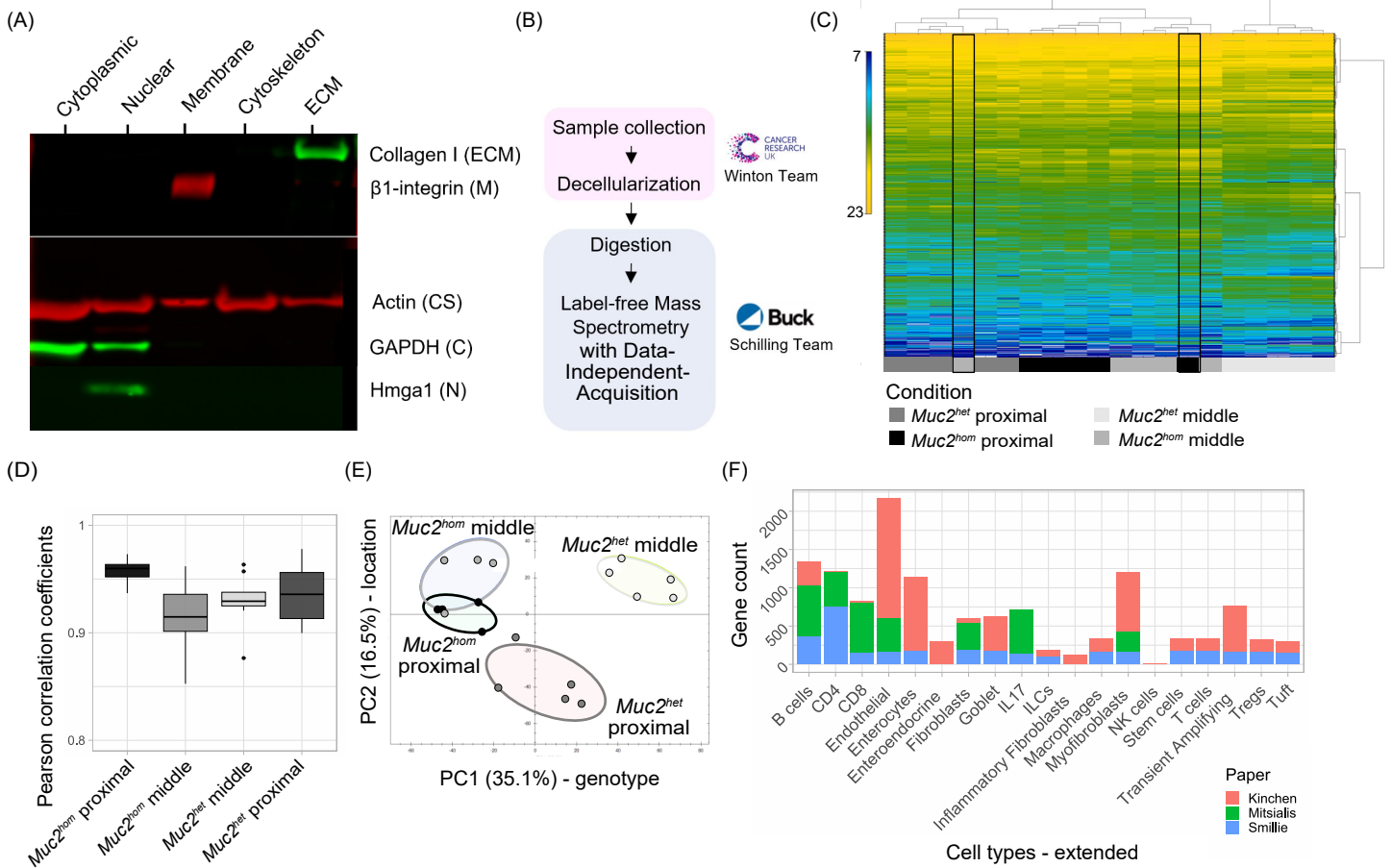
Supplemental Figure 2: Timeline and sample location for ECM proteomics
(A) Timeline sowing age of mice used for histopathological analysis and ECM proteomics. (B) Scheme representing the location of samples used for ECM proteomics and associated level of pathology.

Supplemental Figure 3: Experimental approach to investigate ECM remodelling in colitis
(A) Western blot showing sequential purification of the extracellular tissue fraction. Gapdh was used to probe for removal of cytoplasmic proteins, Hmga1 of nuclear proteins, β1-integrin of membrane proteins and Actin of cytoskeletal proteins. Enrichment for ECM is seen with appearance of a band for Collagen1. (B) Experimental outline for label-free mass-spectrometry on decellularized fractions of $Muc2^{KO}$ mouse colon. (C) Heatmap displaying identified and quantified protein groups for all samples, showing 2 out-layers outlined in black that were removed from the analysis. (D) Box plots representing the distribution of Pearson correlation coefficients for samples in each condition. (E) Principal Component Analysis using the abundance of all quantified proteins between sample groups, where PC1 (x-axis) separates samples based on genotype and PC2 (y-axis) based on location along the colon. (F) Stacked bar plots showing the number of genes present in each extended cell type signatures, coming from each of 3 colon single-cell RNA sequencing published datasets (see details in methods).
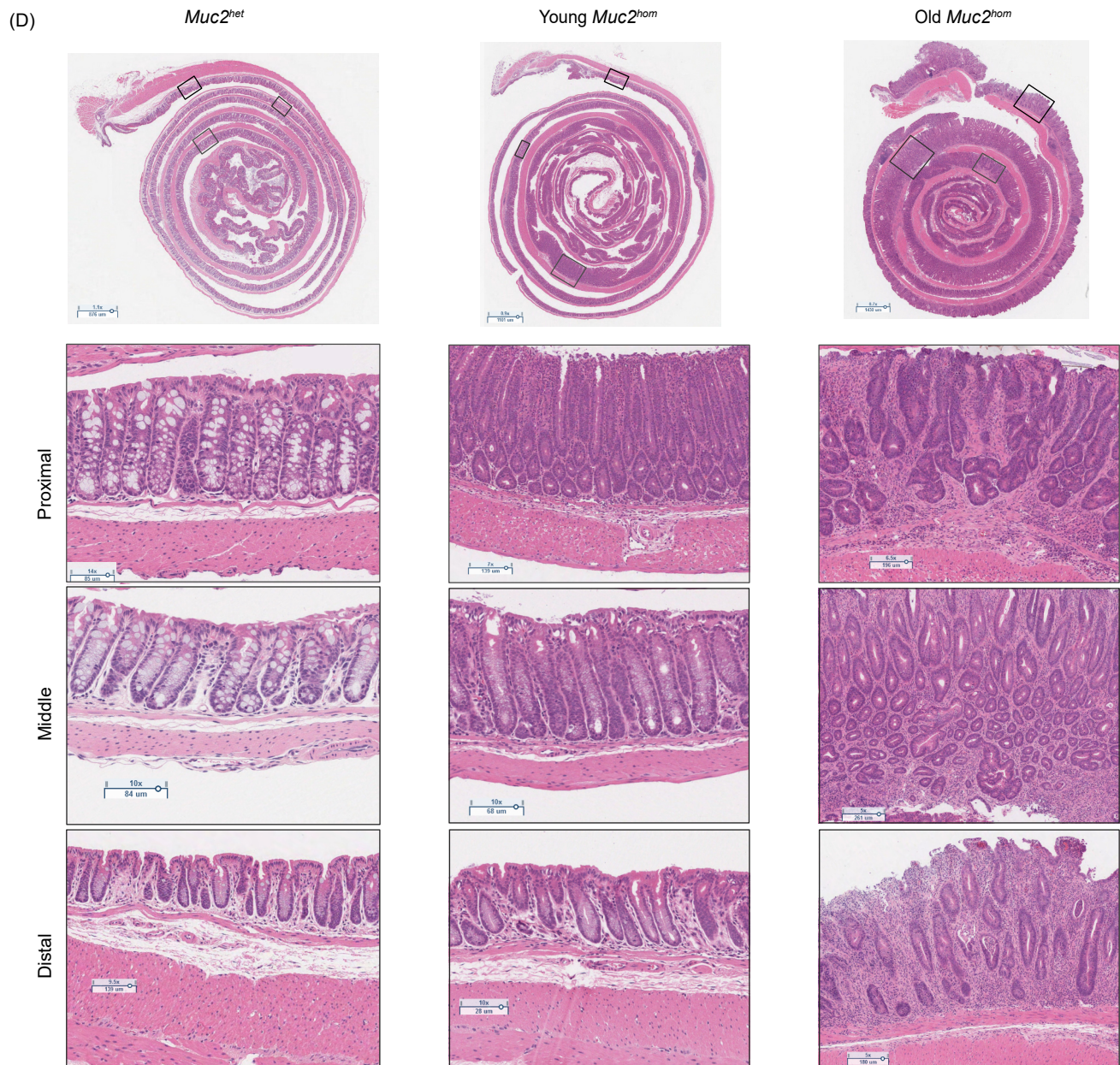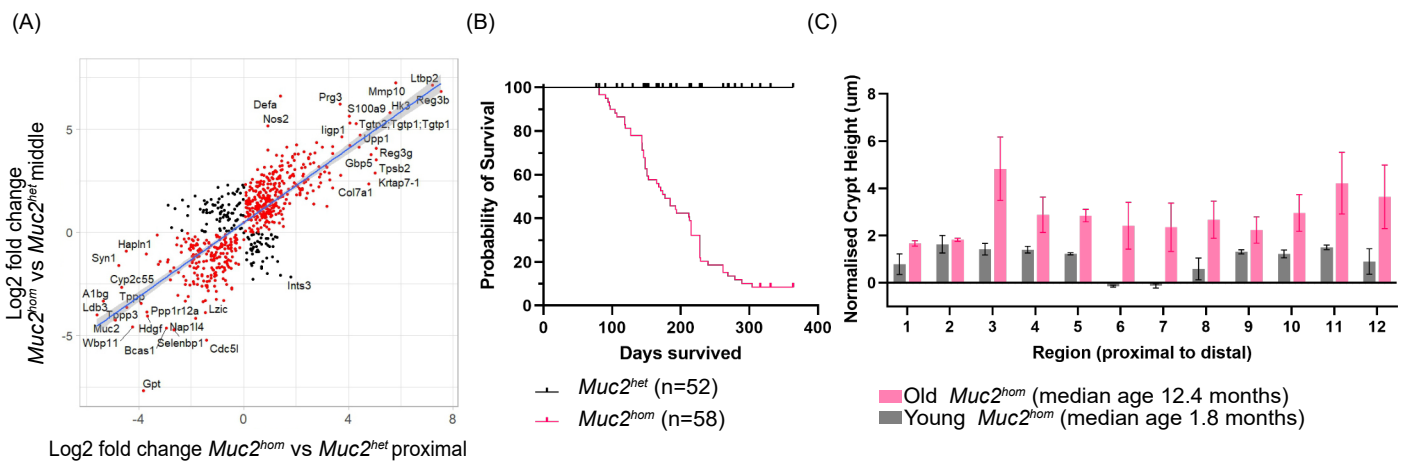
Supplemental Figure 4: *Muc2^{KO}* mice evolve from discrete to pan-colitis with age
(A) Dotplot displaying log2 fold changes for each protein differentially expressed in *Muc2^{hom}* mid- colon (y-axis) and proximal (x-axis) samples compared to location matched *Muc2^{het}* colon sample (Qvalue < 0.05). (B) Survival analysis showing heterogeneity in *Muc2^{hom}* mice survival, from 100 days up to 1-year-old (study endpoint). (C) Bar plot showing crypt height for young (in grey, median age 1.8 month) and old (in pink, median age 12.4 months) *Muc2^{hom}* mice, normalised to age matched *Muc2^{het}* controls (n=3 mice for each condition). (D) Representative H&E pictures showing development of pathology in the proximal colon in young *Muc2^{hom}* mice, which extends to the whole colon length in old *Muc2^{hom}* mice.
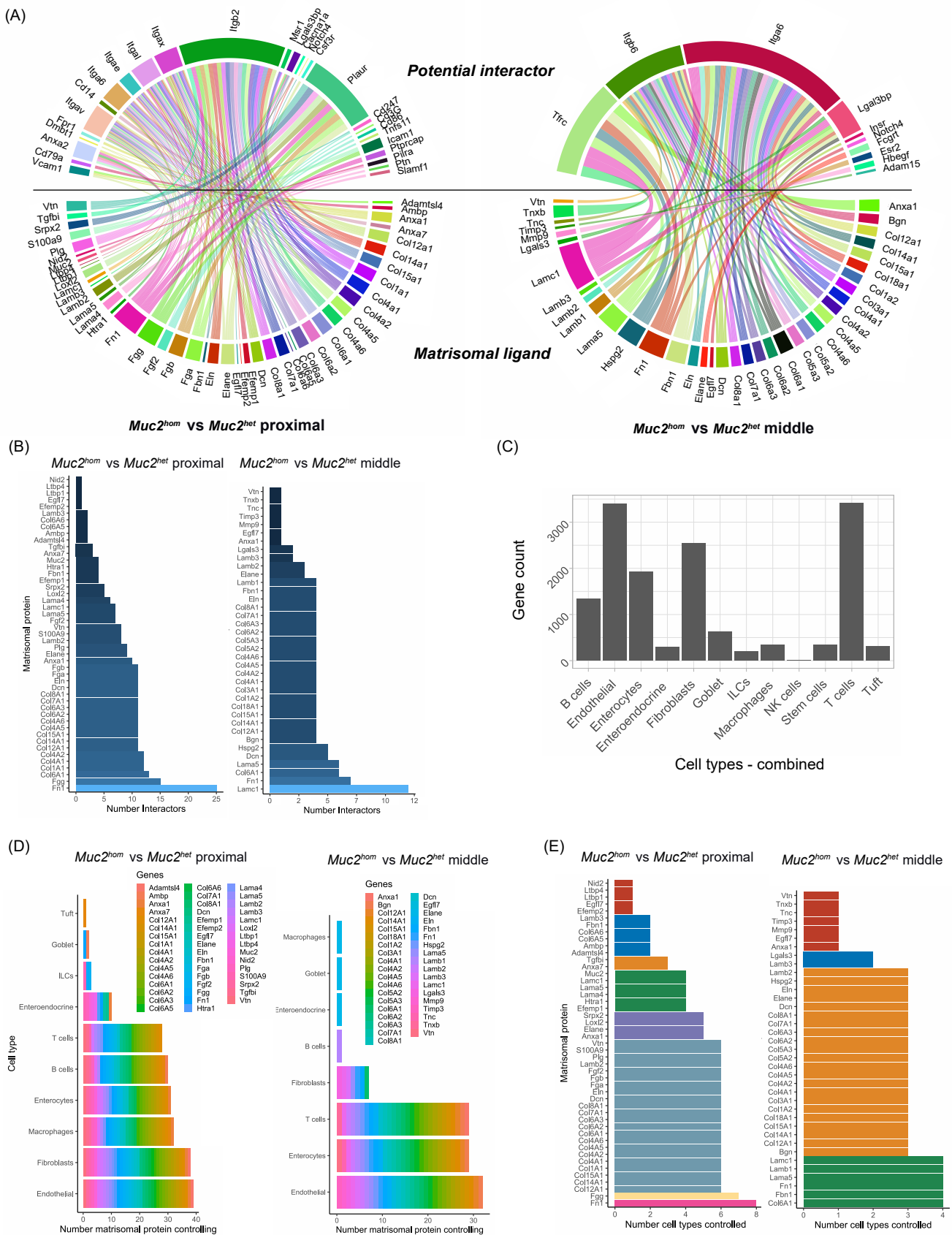
Supplemental Figure 5: Building networks representing matrisomal orchestration of cell type remodelling

(A) Circos plot showing differentially expressed matrisomal ligands (bottom, Qvalue < 0.05) and their known interactors differentially expressed at the transcriptomic level (top, adjusted p value < 0.05), in proximal *Muc2^hom^* (left) and middle *Muc2^hom^* (right) compared to location-matched *Muc2^het^* control samples. (B) Histograms showing the number of receptors differentially expressed at the transcriptomic level (x a-axis, adjusted p value < 0.05), for each matrisomal ligand differentially expressed in the comparison of proximal *Muc2^hom^* vs *Muc2^het^* (left) or middle *Muc2^hom^* vs *Muc2^het^* (right). (C) Bar plot representing the number of genes present in each combined cell type signatures, where discrete cell types were combined under a general cell type (see method). (D) Histogram showing the number (x-axis) and name (figure legend) of matrisomal proteins (as described in B) controlling each cell type present on the y-axis (combined cell type signature). (E) Histogram displaying the number of cell types (combined cell type signature) controlled by each matrisomal protein as described in B.