Cellular and molecular dysregulation of the esophageal epithelium in systemic sclerosis

Matthew Dapas[1,2], Margarette H. Clevenger[3], Hadijat M. Makinde[1,2], Tyler Therron[1], Dustin A. Carlson[3,4], Mary Carns[1], Kathleen Aren[1], Cenfu Wei[3], Lutfiyya N. Muhammad[5], Carrie L. Richardson[1], Parambir S. Dulai[2,3], Monique Hinchcliff[6], John E. Pandolfino[3,4], Harris R. Perlman[1,2*], Deborah R. Winter[1,2*], Marie-Pier Tetreault[2,3*]

Affiliations:

1. Division of Rheumatology, Department of Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL, USA
2. Center for Human Immunobiology, Northwestern University Feinberg School of Medicine, Chicago, IL, USA
3. Division of Gastroenterology & Hepatology, Department of Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL, USA
4. Kenneth C. Griffin Esophageal Center of Northwestern Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL, USA
5. Department of Preventive Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL, USA
6. Section of Rheumatology, Allergy & Immunology, Department of Medicine, Yale School of Medicine, New Haven, CT, USA

*These authors jointly supervised the work

ABSTRACT

Systemic sclerosis (SSc) is a rare autoimmune disease characterized by vasculopathy and fibrosis of the skin and internal organs. Individuals with SSc often suffer from chronic acid reflux and dysphagia due to loss of esophageal motility. However, the pathogenesis of esophageal dysmotility in SSc is poorly understood. To determine whether distinct changes in esophageal epithelial cells contribute to impaired motility in SSc, we investigated the stratified squamous esophageal epithelium from proximal and distal biopsies using single-cell RNA sequencing (n=306,372 cells) in individuals with SSc compared those with gastroesophageal reflux disease (GERD) as well as healthy controls. The proportion of epithelial cells in the apical, superficial compartment of the esophageal epithelium was significantly reduced in SSc (9.4% vs 21.6% in HCs). Differential gene expression in SSc was primarily limited to the superficial compartment (3,572 genes vs. 232 in all other compartments, based on pseudobulk analysis), with significant upregulation of extracellular matrix and keratinization genes. These cellular and molecular changes in SSc were highly correlated with those seen in GERD, indicating they were secondary to reflux; however, their magnitudes were more pronounced in the proximal esophagus, suggesting that esophageal dysmotility leads to greater proximal acid exposure, which may contribute to aspiration. SSc-specific gene dysregulation implicated immunoregulatory pathways likely pertinent to pathogenic mechanisms. Cell type localization and SSc-specific changes were confirmed by spatial molecular imaging. By offering a comprehensive view of transcriptional dysregulation at single-cell resolution in human esophageal epithelial cells in SSc compared to GERD and healthy tissue, this work clarifies the state of epithelial cells in SSc-induced esophageal dysfunction.

47 **INTRODUCTION**

48 Systemic sclerosis (SSc), also known as scleroderma, is an immune-mediated rheumatic disease of
49 unknown etiology that causes fibrosis of the skin and internal organs. Although uncommon (18-26 per
50 100,000 people)[1-4], SSc is one of the deadliest autoimmune disorders[5], with a 10-year survival rate of 45-
51 68% following diagnosis[4,6,7]. Over 90% of individuals with SSc report gastrointestinal dysfunction[8], with
52 esophageal dysmotility being the most common gastrointestinal manifestation[9]. Individuals with SSc and
53 esophageal involvement typically suffer from chronic acid reflux and dysphagia due to loss of esophageal
54 motility and are at much greater risk of esophageal stricture and Barrett's esophagus[8,10,11]. The prevalence
55 of esophageal involvement was nearly twice as high in individuals with SSc who died within 5 years of
56 disease diagnosis[12].

57 There are multiple, interrelated mechanisms that may cause esophageal dysmotility in SSc.
58 Vascular damage and/or neurogenic impairment can lead to smooth muscle atrophy[13], which is the
59 predominant esophageal pathology in autopsies of patients with SSc[14]. Inflammatory signatures[15], absence
60 of anti-centromere antibodies[16], and presence of anti-topoisomerase I antibodies are associated with
61 esophageal pathologies, as well[8]. Complicating research efforts is significant clinical and molecular
62 heterogeneity observed between SSc patients[17-19] and difficulty distinguishing between effects from
63 autoimmune processes and secondary factors such as severe gastroesophageal reflux disease (GERD).
64 Consequently, the pathogenesis of esophageal disease in SSc remains poorly understood.

65 Functional investigations of SSc esophageal involvement have primarily focused on peristalsis or
66 contractions of the muscularis propria[20]. However, the mucosa is no less essential for normal esophageal
67 transport[21,22], and recent studies have suggested that esophageal epithelial cells (EECs) may play a more
68 central role in the pathogenesis of SSc esophageal involvement than previously thought[23]. Mouse models
69 with epithelial-cell-specific knockout of *Fli1*, an Ets transcription factor implicated in SSc pathogenesis[24],
70 spontaneously develop dermal and esophageal fibrosis and interstitial lung disease[25]. Importantly, the
71 resultant lung disease is mediated by T-cell autoimmunity, but fibrosis of the skin and esophagus persist in

72 mice additionally lacking mature T and B cells, suggesting that the fibrosis is principally driven by the
73 epithelia[23,25].

74 Most studies implicating epithelial cells in SSc esophageal dysmotility have been performed in
75 mice or *in vitro*, and molecular analyses in humans have so far been limited to bulk tissues[15]. Furthermore,
76 single-cell gene expression studies performed in other diseases of the esophagus, including squamous cell
77 carcinoma[26,27] and allergic eosinophilic esophagitis[28,29], have identified substantial cellular and molecular
78 changes in EECs compared to healthy tissue, such as significant expansion of non-proliferative suprabasal
79 cells[28,29]. Therefore, to comprehensively investigate the effects of SSc on the esophageal epithelium, we
80 performed single-cell RNA sequencing (scRNA-seq) of esophageal mucosal biopsies from SSc patients
81 with clinically significant esophageal involvement and compared EEC distributions and gene expression
82 signatures to non-SSc individuals with GERD and healthy controls (HCs). We further validated cell type
83 localization and SSc effects using spatial molecular imaging from independent SSc and HC samples. By
84 examining gene expression profiles from human tissue at single-cell resolution, this work sheds essential
85 light on the cellular roots of esophageal dysfunction in SSc by clarifying the pathogenic role of the
86 squamous epithelium, one of the most integral tissues supporting healthy esophageal function.

87

88 **METHODS**

89

90 **Subjects**

91 The study protocol received approval from the Northwestern University Institutional Review Board. The
92 SSc study population were recruited from adult patients aged 18 to 89 diagnosed with SSc according to the
93 2013 ACR/EULAR Classification Criteria, who visited the Northwestern Medicine Esophageal Center
94 between 2020 and 2021for esophageal symptom evaluation and motility testing, as described previously[20].
95 Patients with non-SSc GERD were recruited at the first clinic visit following positive wireless pH testing[29].
96 Patients with technically inadequate panometry or manometry studies, previous foregut surgeries (including
97 prior pneumatic dilation), candida esophagitis, history of esophageal cancer, or mechanical obstructions in

98 the esophagus such as esophageal stricture, eosinophilic esophagitis, severe reflux esophagitis (Los Angeles
99 classification C or D), or hiatal hernias larger than 5 cm were excluded. HCs met asymptomatic criteria
100 including lack of esophageal symptoms, no history of alcohol dependency or tobacco use, body mass index
101 less than 30kg/m^2, and no current antacid or proton pump inhibitor treatment[29]. The study protocol received
102 approval from the Northwestern University Institutional Review Board.
103
104 **Symptom evaluation and motility testing**
105 Esophageal function testing included functional lumen imaging probe (FLIP) panometry during sedated
106 endoscopy, high-resolution manometry (HRM), and in some subjects, 24-hour pH-impedance or timed-
107 barium esophagogram, which were completed and interpreted as previously described[30-33]. HRM
108 classifications were based on the application of the Chicago Classification v4.0 to 10 supine and 5 upright
109 test swallows[33]. FLIP panometry classifications were based on previous evaluation of asymptomatic
110 volunteers and patients[34-36]. During endoscopy (same encounter as FLIP), esophageal mucosal biopsies
111 were obtained at approximately 5-cm ("distal esophagus") and 15-cm ("proximal esophagus") proximal to
112 the squamocolumnar junction using Boston Scientific Radial Jaw 4 standard capacity forceps. Acid
113 exposure was measured as the percentage of time with esophageal pH<4 over a monitoring period of 24
114 hours. Most patients also completed validated symptom-questionnaires (patient-reported outcome
115 measures) on the day of esophageal motility testing with FLIP or HRM, including the Brief Esophageal
116 Dysphagia Questionnaire (BEDQ), the Gastroesophageal Reflux Disease Questionnaire (GerdQ), and the
117 Northwestern esophageal quality of life (NEQOL) survey.[37-39]. SSc disease duration was measured from
118 onset of the first non-Raynaud symptom.
119
120 **Single-cell RNA-sequencing sample processing and sequencing**
121 Esophageal mucosal biopsies from proximal and distal esophagus were processed for scRNA-seq as
122 described in Clevenger et al[29]. Briefly, following tissue digestion and filtering, cell suspensions met 85%
123 minimum viability via Cellometer Auto2000 (Nexcelom Bioscience). Cells were loaded into a Chromium

124 iX controller (10X Genomics) on a Chromium Next GEM Chip G (10X Genomics) to capture
125 approximately 10,000 cells per sample and were processed for encapsulation following the manufacturer's
126 protocol. Cell barcoding and library construction were performed using the 10X Genomics Chromium Next
127 GEM Single Cell 3' Reagents Kits v3.1 and Dual Index Kit TT Set A according to the manufacturer's
128 protocol. Sequencing was performed by the NUSeq Core using the Illumina NovaSeq 6000. Paired-end
129 reads consisting of a 28-base-pair read for cell barcodes and unique molecular identifiers and a 90-base-
130 pair read for transcripts were aligned to the GRCh38 reference transcriptome using Cell Ranger v6.1.2.
131

132 **Single-cell sample and cell quality control**

133 Samples with high ambient RNA (<70% of sequencing reads mapped to cells) were removed from
134 consideration. Gene expression counts generated by Cell Ranger were analyzed using Seurat v4.3.0[40] in R
135 v4.3.0. Cells with low gene diversity (<200 unique genes), low unique molecular identifier counts (<500),
136 or excessive counts (>100,000) were removed from consideration. Cells with high proportions of
137 mitochondrial DNA reads, indicative of apoptosis or lysis, were removed using the bivariate regression
138 models implemented in the miQC package[41]. Cells were filtered if their proportion of reads mapped to
139 mitochondrial DNA exceeded 0.05 and their posterior probability of belonging to a compromised cell
140 distribution was greater than 0.75 according to miQC. Doublets, in which more than one cell is clumped
141 together in a single droplet, were predicted and removed using scDblFinder[42] with an expected, additive
142 doublet rate of 0.01 per thousand cells with standard deviation of 0.01.
143

144 **Sample integration and cell type annotation**

145 Individual scRNA-seq expression profiles were normalized and integrated on 3,000 variable features using
146 Seurat's SCTransform procedure[43] with v2 regularization[44] and canonical correlation analysis dimension
147 reduction. Integration anchors were determined using a reference consisting of three pairs of proximal and
148 distal samples from HCs. Principal component (PC) analysis was performed on the transformed and
149 integrated expression counts. Uniform manifold approximation and projection (UMAP)[45] embeddings were

150 calculated on the top 40 PCs with n=200 neighbors. Cell clustering was performed using Seurat's
151 modularity-based clustering on a shared nearest neighbor (SNN) graph with k=25 on the top 40 PCs.
152 Clusters were annotated according to the expression of established cell-specific gene markers
153 (**Supplemental Figure 1**). For annotation of smaller cell clusters, we first filtered out the epithelial,
154 myeloid, lymphoid, and endothelial cell clusters and then recomputed PCs and re-clustered the subset (PCs
155 = 25, k=15).
156
157 **Epithelial cell characterization**
158 After isolating the subset of epithelial cell clusters, we further filtered cells with disproportionately low
159 read counts (≤3,500; **Supplemental Figure 2, A-E**). The samples were then re-integrated using the
160 procedure described above. UMAP embeddings were calculated on the top 35 PCs with n=100 neighbors.
161 Epithelial compartments (basal, proliferating basal, proliferating suprabasal, suprabasal, and superficial)
162 were identified according to expression of established compartment-specific and cell cycle markers
163 (**Supplemental Figure 2F**)[28,29,46-48]. For the basal, suprabasal, and superficial compartments, we computed
164 composite gene signature scores using Seurat's AddModuleScore() function (**Supplemental Figure 2G**).
165 Cells were initially clustered on an SNN graph with k=50 with resolution=0.25 (**Supplemental Figure**
166 **2H**), but because the modularity optimization clustering implemented in Seurat grouped proliferating cells
167 across different epithelial compartments and inconsistently distinguished suprabasal and superficial cells,
168 we applied K-means clustering on the expression of the compartment-specific markers (**Supplemental**
169 **Figure 2F**). We performed separate K=2 K-means clustering on 1) the Seurat clusters containing basal and
170 early suprabasal cells (clusters 1-5) to distinguish basal from suprabasal compartments and 2) the Seurat
171 clusters containing late suprabasal and superficial cells (clusters 6-8) to demarcate suprabasal and
172 superficial compartments (**Supplemental Figure 2, I-J**). Proliferating epithelial cells were identified using
173 Seurat's CellCycleScoring() function. Cells predicted as being in S or G2/M phases based on the relative
174 expression of corresponding markers[49] were labeled as proliferating cells (**Supplemental Figure 2K**). We

175 tested for statistical differences in proportions of epithelial cell populations between sample groups using
176 MASC[50].
177
178 **Gene expression analysis**
179 To identify dysregulated pathways in SSc, we first calculated differential gene expression between
180 conditions (SSc, non-SSc GERD, HCs) within each epithelial compartment, considering each cell as a
181 sample. Differential gene expression was calculated using the FindMarkers() function in Seurat with the
182 Wilcoxon rank-sum test on log-normalized RNA counts for genes expressed in at least 1% of cells by
183 compartment. We adjusted for multiple testing using a Bonferroni correction accounting for the number of
184 genes tested. Genes with an adjusted $P<0.05$ and absolute $\log_2$ fold change >0.1 were considered
185 significantly differentially expressed. [51,52] Genes were then ranked by the magnitude and statistical certainty
186 of their expression differences:
187
188 $$Weight = -\log_{10}(p) * abs(\log_2[FC])$$
189
190 Gene set enrichment was calculated using 3,795 canonical pathway gene sets from the Human Molecular
191 Signatures Database[53] using gene set enrichment analysis (GSEA)[51,52]. Statistical gene set enrichment was
192 calculated using one-way positive enrichment with 100,000 permutations on pathways with at least 10
193 genes. Mitochondrial and ribosomal genes were excluded from the GSEA.
194 For pseudobulk differential expression analysis we aggregated RNA counts per sample within
195 epithelial compartments. Genes with at least one count per million detected in at least 75% of samples in at
196 least one condition per epithelial compartment were analyzed. To mitigate outlier-driven signals, genes
197 with log-transformed expression greater than 3 standard deviations from the mean in exactly one individual
198 were excluded. Differential gene expression was evaluated using edgeR v3.42.4[54-56]. Gene expression was
199 modeled using a quasi-likelihood negative binomial generalized linear model, with global negative-
200 binomial trended dispersion estimated across all samples and gene-specific quasi-likelihood dispersions

201 estimated within each epithelial compartment. Differences in gene expression across conditions were
202 assessed using quasi-likelihood F-tests with false-discovery rate (FDR) adjustment. Differences in the
203 proportions of genes that were significantly differentially expressed between groups were assessed using a
204 chi-square test. Differences in gene expression profiles between biopsy locations were determined by
205 measuring intra-individual correlations for the 2,000 most variable genes at the single-cell level aggregated
206 by sample, epithelial compartment, and biopsy location.

207 Transcription factor enrichment analyses were conducted on the pseudobulk results using the
208 enricher() function from clusterProfiler v4.8.3[57]. We used consensus results from ENCODE[58] and ChEA[59],
209 as curated by the Ma'ayan Lab[60] and available at http://amp.pharm.mssm.edu/Enrichr/[61]to test for the
210 enrichment of transcription factor target genes from among genes that were differentially expressed
211 between SSc, GERD, and HCs with Bonferroni-adjusted $p<0.05$. Background gene sets were comprised of
212 the genes included in the pseudobulk differential expression analyses for each epithelial compartment.
213 Transcription factors absent from the background gene set were removed as candidate transcription factors.
214 Separate enrichment tests were performed for all differentially expressed genes and those that were
215 upregulated or downregulated in SSc. Enrichment results were adjusted using a Bonferroni correction
216 accounting for the number of transcription factors with more than 2 genes that were significantly
217 differentially expressed.

218

219 **Pseudobulk sample permutation**

220 Because of the sample size imbalance by disease state, which inherently yields different statistical power
221 for different tests across conditions, we permuted differential expression testing with equivalent sample
222 sizes to better evaluate the relative extent of dysregulation by EEC compartment and biopsy location
223 between SSc and GERD. We permuted all possible combinations of SSc samples with sample size n=4 for
224 both the proximal and distal regions and used the median statistics in the SSc vs. HC tests to more directly
225 compare the results with those observed in GERD vs. HCs.

226

227 **Spatial imaging sample preparation**

228 Tissue sections were mounted on Leica Biosystems Apex BOND Superior Adhesive Slides (cat# 3800040)
229 and subsequently baked overnight at 60 °C, followed by preparation for in-situ hybridization (ISH) by
230 deparaffinization in CitriSolv (2x 5min), 100% ethanol (2x 5min) and heat-induced epitope retrieval
231 (HIER) at 100 °C for 15 min using low pH citrate buffer (NanoString/Bruker). After HIER, tissue sections
232 were digested with Proteinase K (2 µg/ml) diluted in PBS at 40 °C for 30 min. Tissue sections were then
233 washed twice with diethyl pyrocarbonate (DEPC)-treated water (DEPC H2O) and incubated in 0.0015%
234 fiducials (NanoString/Bruker) in 2X saline sodium citrate, 0.001% Tween-20 (SSCT solution) for 5 min at
235 room temperature in the dark. Excess fiducials were rinsed from the slides with 1X PBS, then tissue sections
236 were fixed with 10% neutral buffered formalin (NBF) for 5 min at room temperature. Fixed samples were
237 rinsed twice with Tris-glycine buffer (0.1 M glycine, 0.1 M Tris-base in DEPC H2O) and once with 1X
238 PBS for 5 min each before blocking with 100 mM N-succinimidyl (acetylthio)acetate (NHS-acetate,
239 Thermo Fisher Scientific) in NHS-acetate buffer (0.1 M NaP, 0.1% Tween pH 8 in DEPC H2O) for 15 min
240 at room temperature. The sections were then rinsed with 2X saline sodium citrate (SSC) for 5 min and an
241 Adhesive Hybridization Chamber (NanoString/Bruker) was placed over the tissue. ISH probes were
242 prepared by incubation at 95 °C for 2 min and submerged in ice, and the ISH probe mix was pipetted into
243 the hybridization chamber. The chamber was sealed to prevent evaporation, and hybridization was
244 performed at 37 °C for 18 hrs. Tissue sections were washed twice in 50% formamide (Ambion) in 2X SSC
245 at 37 °C for 25 min, washed twice with 2X SSC for 2 min at room temperature, and blocked with 100 mM
246 NHS-acetate in the dark for 15 min. In preparation for loading onto the spatial imaging instrument, a flow
247 cell was affixed to the slide.

248

249 **Spatial molecular imaging and data processing**

250 Single-cell spatial transcriptomic analysis was performed using the CosMx Spatial Molecular Imager
251 (NanoString/Bruker), featuring the 6k Human Discovery Panel gene list, which consists of 6,175 RNA
252 targets designed to provide coverage across all major biological pathways. Spatial fields of view (FOVs)

253 with average counts per cell <100 were removed from consideration. Cells with low RNA counts (<20),
254 high proportions of counts from negative probes (>0.1), extreme size (p<0.01 by Grubb's test), or more
255 detected genes than counts were removed. Gene probes with expression levels below median background
256 noise, as determined by negative probe detection, were also removed. Counts were log normalized using
257 Seurat[40]. Cell segmentation errors were predicted using scDblFinder[42]. To annotate the cells, we performed
258 label transfer from the single-cell dataset with SingleR[62] using default parameters on the intersect of
259 quantified genes. Gene expression differences between groups in the spatial data were evaluated using
260 Wilcoxon rank-sum tests.
261
262 **Clinical phenotype associations**
263 To test for associations with categorical measures of esophageal function (FLIP, HRM, and a combined
264 FLIP and HRM with neuromyogenic [NM] classification[63]), we performed ordinal logistic regression that
265 modeled likelihoods of increasing clinical severity. For this analysis, we excluded SSc patients with
266 achalasia or esophagogastric junction outflow obstruction (EGJOO). For FLIP we combined the
267 "borderline/diminished" and "impaired/disordered" phenotypes. For NM we combined the Stage I and
268 Stage II ineffective motility phenotypes. The resultant ordering for the ordinal regression was "normal" <
269 "ineffective" < "absent" for HRM; "normal" < "borderline/diminished & impaired/disordered" < "absent"
270 for FLIP; and "normal" < "stages I & II – ineffective" < "stage III – absent" for NM.
271 For association testing with quantitative clinical measures in SSc, we first performed a principal
272 components analysis (PCA) on a set of nine quantitative clinical traits: HRM mean distal contractile interval
273 (DCI), HRM basal esophagogastric junction (EGJ) pressure, HRM EGJ contractile index, HRM median
274 integrated relaxation pressure (IRP), FLIP intra-balloon pressure at 60ml, FLIP EGJ distensibility index,
275 the GerdQ impact score, the BEDQ score, and the NEQOL score. Missing values were imputed with the
276 mean across all patients. We then modeled associations with the first two PCs using linear regression. Gene
277 expression was represented using log-transformed counts.
278

279

280 **RESULTS**

281 Whole tissue scRNA-seq was performed on paired proximal and distal mucosal biopsies from ten patients
282 with SSc, four non-SSc patients with GERD, and six HCs. All participants but one in each group were
283 female (**Table 1**). SSc patients were significantly older than HCs (54.9 vs 27.0 years; P<0.001). Among
284 SSc patients, six had the limited cutaneous subtype, three had the diffuse cutaneous subtype, and one had
285 sine scleroderma with no cutaneous symptoms (**Supplemental Table 1**). Mean SSc disease duration at the
286 time of biopsy was 132 months. All SSc patients had positive serum antinuclear antibodies and impaired
287 esophageal motility. Following sample-level quality control (**Supplemental Table 2**), 39 samples were
288 retained for analysis, including 19 from the proximal esophagus and 20 from the distal esophagus. Four
289 additional sets of proximal and distal biopsies were collected for spatial molecular imaging from two SSc
290 patients with diminished esophageal contractility and two healthy controls (**Supplemental Table 3**).

291

292 **Characterization of esophageal mucosal cell populations**

293 A total of 306,372 esophageal cells (7,856 mean cells per sample) were retained for cell clustering and gene
294 expression analysis. Each cell type was identified according to the expression of established cell-specific
295 gene markers including *KRT6A, KRT13,* and *KRT15* for epithelial cells (**Figure 1, A-D; Supplemental**
296 **Table 4**; **Supplemental Figure 1**). Epithelial cells comprised the bulk of the sample (n=264,858; 86.45%),
297 followed by lymphoid cells (n=25,129; 8.20%), myeloid cells (n=10,019; 3.27%), endothelial cells
298 (n=4,027; 1.31%), and all other cell types (n=2,339; 0.76%). We observed heterogeneity in non-epithelial
299 cell type proportions by condition and biopsy location (**Figure 1E**) but focused on the epithelium for the
300 remainder of the study.

301

302 **Significant loss of superficial EECs in SSc**

303 Following isolation of the epithelial cell cluster and additional quality control (**Supplemental Figure 2, A-**
304 **E**), we characterized the remaining 230,720 EECs according to their differentiation and cell cycling states

305 (**Figure 2A**). We classified the EECs into five primary compartments: basal (n=55,818; 24%), proliferating
306 basal (n=36,439; 16%), proliferating suprabasal (n=25,314; 11%), suprabasal (n=82,283; 36%), and
307 superficial (n=30,866; 13%) based on relative expression of canonical epithelial genes and cell cycle
308 markers (**Figure 2, B-C; Supplemental Figure 2, F-K**). Spatial localization of the annotated single-cell
309 clusters was confirmed using spatial molecular imaging (**Figure 2D, Supplemental Figure 3**). Roughly
310 40% of basal cells and 24% of suprabasal cells were proliferating, and the fractions of cells that were
311 proliferating did not differ by disease (**Figure 2E**). Proportions of basal and suprabasal cells were not
312 significantly different across conditions, but there were significantly fewer proximal superficial cells in SSc
313 compared to HCs (μ_{SSc}=0.10, μ_{HC}=0.21, P=0.003, **Figure 2E**).
314
315 **EEC gene dysregulation predominantly in superficial cells, highly correlated in SSc and GERD**
316 Most differentially expressed genes (DEGs) were only differentially expressed in one epithelial
317 compartment (**Supplemental Figure 4A**), and differential gene expression between conditions was most
318 pronounced in superficial EECs (**Figure 3A**). At the single-cell level, 674 and 434 genes were significantly
319 differentially expressed between SSc and HCs in the superficial compartment in the proximal and distal
320 regions, respectively, compared to 171 proximal and 157 distal non-proliferating basal DEGs and 172
321 proximal and 377 distal non-proliferating suprabasal DEGs (**Supplemental Tables 5-6**). To more robustly
322 characterize gene expression differences between conditions[64], we performed pseudobulk differential
323 expression analysis by aggregating single-cell RNA counts per sample within EEC compartments. Overall,
324 pseudobulk gene expression patterns clustered by compartment rather than by condition or biopsy location
325 (**Supplemental Figure 4B**). Significant differences in gene expression between conditions were largely
326 limited to superficial cells, suggesting that the apical cell layers of the esophageal epithelium are most
327 affected by SSc (**Supplemental Figure 4, C-D**). There were 3,572 genes differentially expressed between
328 SSc and HCs with FDR q<0.05 in the superficial compartment, compared to just 232 total in all other
329 compartments (**Supplemental Table 7**).

330 The differential expression observed in SSc was strongly correlated with the changes seen in
331 GERD, with many of the same genes significantly dysregulated in both SSc and GERD compared to HCs.
332 Overall correlations in pseudobulk $\log_2$ fold change between SSc vs. HCs and GERD vs. HCs were
333 consistently strong, ranging between r=0.51-0.62 across EEC compartments (**Supplemental Table 7**).
334 Depending on the compartment and biopsy location, 22-44% of single-cell DEGs in SSc or GERD were
335 differentially expressed in both conditions (**Figure 3A**; **Supplemental Table 5**). Consequently, the single-
336 cell DEGs in SSc and GERD were significantly enriched for some of the same pathways, particularly those
337 related to the extracellular matrix and keratinization (**Figure 3B**). These included matrisome (suprabasal,
338 distal; superficial, distal) and matrisome-associated genes (suprabasal, distal); cornified envelope formation
339 (suprabasal, distal; superficial, distal), and keratinization genes (suprabasal, distal; superficial, distal);
340 developmental biology-related genes (suprabasal, distal; superficial, distal). The leading edges of these
341 significant pathways, which comprised the genes driving their enrichment, included shared protein families,
342 including serine protease inhibitors (serpins), keratins, small proline-rich proteins (SPRRs) and S100
343 proteins (**Supplemental Table 8**). Among the most enriched pathways specific to SSc were those related
344 to the regulation of metal and immune homeostasis. In particular, innate immune system genes were
345 uniquely enriched in SSc compared to HCs in superficial EECs of the proximal esophagus.

346 Overall gene expression levels were highly similar between esophageal biopsy locations, with mean
347 intra-individual correlations grouped by condition and epithelial compartment ranging from r=0.95 to
348 r=0.99 for the 2,000 most variable genes (**Figure 3C**; **Supplemental Table 9**). However, we did observe
349 some notable differences in differential gene expression patterns by biopsy location. For example, single-
350 cell differential expression in the suprabasal compartment was significantly greater in the distal esophagus
351 than in the proximal esophagus for both SSc (172 proximal vs. 377 distal, $p=2.5\times10^{-20}$) and GERD (159
352 proximal vs. 434 distal, $p=1.1\times10^{-32}$; **Figure 3A; Supplemental Figure 4A**). Furthermore, at the single-
353 cell level, there were more unique DEGs in SSc vs. HCs in the proximal esophagus (897 proximal vs. 806
354 distal, $p=7.0\times10^{-2}$), but the opposite was true in GERD vs. HCs (910 proximal vs. 1,075 distal, $p=1.6\times10^{-5}$;

355 **Figure 3A**). Similarly, within the superficial compartment, where most differential expression was
356 observed, GERD exhibited greater relative changes in pseudobulk gene expression in the distal esophagus
357 than in the proximal region, as indicated by the decreased slope between fold-changes (m=0.44 vs. 0.87;
358 **Figure 3D; Supplemental Table 10**). The superficial compartment also featured the greatest range and
359 lowest overall values of intra-individual gene expression correlations between biopsy locations, particularly
360 in SSc (**Figure 3C**). These differences by biopsy location suggest that the relative disease effect on the
361 proximal esophagus compared to the distal esophagus is greater within SSc than within GERD.

362 More genes were significantly differentially expressed in SSc than in GERD in the pseudobulk
363 analysis, but in terms of relative fold change vs. HCs, we observed greater expression changes in GERD
364 (**Figure 3D**). To determine whether the greater number of SSc samples were driving this discrepancy, we
365 performed all possible sample permutations with equal numbers of SSc and GERD samples (n=4) and
366 repeated differential gene expression testing to more directly compare the relative number of DEGs in both
367 conditions. In superficial cells from the proximal esophagus, there were more DEGs in SSc in 60% of
368 permutations, but the relative FC differences were greater in GERD in 61% of permutations (**Supplemental**
369 **Figure 4, E-F; Supplemental Table 11**). In the distal region, there were more DEGs in GERD in 62% of
370 permutations, and the relative FC differences were greater in GERD in all permutations. The correlations
371 between SSc and GERD in terms of relative gene expression changes were consistently much higher in the
372 proximal region permutations than in the distal region (**Supplemental Figure 4G**). The permutation results
373 confirmed that the differences in gene expression observed in SSc relative to GERD were not simply due
374 to differences in sample size. In summary, the superficial epithelium is most altered in SSc, exhibiting
375 similar transcriptional changes to GERD, but with relatively greater effect in the proximal esophagus
376 compared with distal.

377

378 **SSc-specific gene dysregulation points to immunoregulatory pathways**

379 Among the most upregulated SSc-specific genes (**Figure 3D; Supplemental Figure 4H**) were genes
380 related to inflammation (*PTGES*, *MFGE8*)[65,66], innate immune response (*FCGBP, BST2*, *CD44*,

381 *APOBEC3A*)[67-70], immune cell migration (*C10orf99*, *LTB4R*, *ACKR3*)[71-73], antigen presentation (*HLA-B*,
382 *CD74*, *TAP1*, *PSMB8*, *PSMB9*)[74], natural killer cell activation (*CLEC2B*)[75], and fibroproliferation (*SGK1*,
383 *HBEGF*)[76,77]. Four of the top five and eight of the top 20 most downregulated SSc-specific genes by fold
384 change in superficial EECs in the proximal esophagus were metallothioneins (*MT1A*, *MT1E*, *MT1F*, *MT1G*,
385 *MT1H*, *MT1M*, *MT1X*, *MT2A*). The metallothioneins comprised the bulk of three pathways uniquely
386 enriched in SSc with P<0.001: zinc homeostasis, copper homeostasis, and cellular responses to stimuli
387 (**Figure 3B**). The most down-regulated SSc-specific gene across both the proximal and distal esophagus
388 was the *H19* lncRNA. Interestingly, the most down-regulated gene by fold change in GERD, *MUC22*, was
389 not differentially expressed in SSc. One gene, *SLC8A1-AS1*, was significantly differentially expressed in
390 both GERD and SSc, but in opposite directions (**Figure 3D; Supplemental Figure 4H; Supplemental**
391 **Table 10**).
392
393 **Transcription factors enriched in proximal SSc esophagus**
394 We next examined whether DEGs in SSc superficial cells were significantly enriched for targets of specific
395 transcription factors. Among genes differentially expressed in SSc compared to GERD, the target genes of
396 IRF1 were significantly enriched (**Supplemental Table 12**). IRF1 targets were significantly enriched
397 among all proximal superficial DEGs (p_{adj}=0.021) and among only those that were downregulated
398 (p_{adj}=0.0024). However, *IRF1* was not itself differentially expressed in superficial cells in SSc compared to
399 GERD. Compared to HCs, three transcription factors were significantly enriched in SSc in the proximal
400 esophagus, including MYC ($p_{adj}=2.1\times10^{-7}$) and E2F4 ($p_{adj}=5.5\times10^{-5}$) for upregulated DEGs and NFE2L2
401 for downregulated DEGs ($p_{adj}=7.6\times10^{-4}$). *MYC* and *E2F4* were also themselves differentially expressed in
402 SSc compared to HCs in superficial cells ($p_{adj}=7.8\times10^{-5}$, $p_{adj}=9.8\times10^{-4}$, respectively). MYC was the most
403 enriched transcription factor in genes upregulated in GERD compared to HCs, but the enrichment was not
404 significant after adjusting for multiple testing (p_{adj}=0.11). No transcription factors were significantly
405 enriched in any pairwise comparison in the distal esophagus.

406

407 **Fewer metallothioneins and metallothionein-expressing cells in superficial EECs in SSc**

408 With SSc-specific dysregulation most prominent in the superficial cells of the proximal esophagus, we
409 performed sub-clustering of these cells to further examine corresponding expression patterns. We identified
410 five distinct clusters (**Figure 4A**) with unique expression marker patterns (**Figure 4B, Supplemental**
411 **Figure 5A, Supplemental Table 13**) and varying degrees of differentiation (**Figure 4C**). We then
412 performed label transfer onto the spatially mapped cells to confirm the physical distribution of these clusters
413 within the esophageal epithelium. Clusters 1 and 3 represented the apical, terminally differentiated
414 epithelial layers, as evidenced by their *FLG* expression[78-80] and spatial localization (**Figure 4D**), but cluster
415 3 had uniquely high expression of metallothioneins (**Figure 4B, Supplemental Figure 5A**). Cluster 2 was
416 distinguished by its relatively high *CTSV* expression, while clusters 4 and 5 were the least differentiated of
417 the superficial cells and were distinguished by their relative expression of *GJB6* and *FGFBP1*, respectively.
418 Cluster 3 was less abundant in SSc compared to HCs with $P<0.05$, while no other clusters demonstrated
419 significantly different abundances between SSc and HCs (**Figure 4E**). Metallothionein expression was also
420 significantly lower in SSc compared to HCs or GERD in each cluster, except for cluster 5 ($P<5\times10^{-14}$,
421 **Figure 4, F-G**). These observations were reproduced in the spatial imaging data, where cluster 3 was half
422 as abundant in SSc (14.5%) as in HCs (31.6%, **Figure 4D**) and metallothionein expression was likewise
423 significantly lower in SSc compared to HCs in clusters 1-4 ($P<1.5\times10^{-7}$, **Figure 4, H-I**). Moreover,
424 metallothionein expression was not significantly different between SSc and HCs in distal superficial cells
425 in the spatial imaging data (**Figure 4I**).

426 Regarding the expression of the significantly enriched transcription factors (IRF1, MYC, E2F4,
427 and NFE2L2; **Supplemental Table 12**) and their target genes, we observed several patterns (**Supplemental**
428 **Figure 5, B-E**). For *MYC*, *E2F4*, and *NFE2L2*, each gene's expression was highest on average in cluster 5,
429 the least differentiated superficial cluster (**Figure 4C**), and by condition was highest in SSc, followed by
430 GERD. Aggregate expression of target genes was positively correlated with transcription factor expression

431 for *MYC* (r=0.42) and *E2F4* (r=0.24) and negatively correlated for *IRF1* (r=-0.20) and *NFE2L2* (r=-0.18).
432 Correspondingly, cluster 5 featured the highest target gene expression for *MYC* and *E2F4* and lowest for
433 *NFE2L2*, and target gene expression differences relative to HCs were likewise correlated in SSc and GERD
434 but were more pronounced in SSc. Notably, several of the immune-related genes uniquely upregulated in
435 SSc (**Figure 3D; Supplemental Figure 4H**) displayed a similar expression pattern within superficial cells
436 as *MYC* and *E2F4*, with expression unique to cluster 5 (**Supplemental Figure 5F**, **Supplemental Table**
437 **13**). For *IRF1*, however, expression of the transcription factor and its target genes changes differed more in
438 GERD than in SSc relative to HCs.
439
440 ***FLI1* not expressed in human EECs**
441 Deletion of the *Fli1* gene in epithelial cells in mice has been shown to recapitulate histological and
442 molecular features of esophageal involvement in SSc[25]. Therefore, we investigated the expression of *FLI1*
443 in our human esophageal samples. The expression of *FLI1* in human EECs was negligible and did not meet
444 the minimum expression thresholds for inclusion in our differential gene expression analysis. We did
445 observe low *FLI1* expression levels in endothelial cells (**Supplemental Figure 6A**), but the expression
446 differences between SSc, HCs, and GERD was not significant (**Supplemental Figure 6B**). There were
447 many FLI1 downstream targets among the proximal, superficial DEGs between SSc and HCs and seven
448 between SSc and GERD (*ARPC2*, *RAB24*, *MTPN*, *PPIF* upregulated; *CEBPZ*, *TRMT10C*, *TBP*
449 downregulated); however, the proportion of these genes among all DEGs was not different than what would
450 be expected by chance (P>0.05), and the average expression of FLI1 targets was higher in GERD and lower
451 in HCs compared to the expression in SSc (**Supplemental Figure 6C**).
452
453 **Reduction of outer EECs and metallothionein expression correlates with loss of contractility in SSc**
454 We next evaluated whether the cellular and molecular changes we observed between SSc, GERD, and HCs
455 were correlated with specific clinical measures of esophageal dysfunction (**Table 1**). To more efficiently
456 model the covariance structure among quantitative, functional esophageal measures in SSc (**Figure 5A**),

457 we performed PCA on a set of nine clinical metrics and tested for correlations with the first two PCs (**Figure**
458 **5B**). The first PC explained 45.5% of variance and was most correlated with FLIP intra-balloon pressure at
459 60ml (r=0.94), The second PC explained 22.6% of variance and was most correlated with HRM basal EGJ
460 pressure (r=0.74).

461 We did not observe any significant associations between epithelial cell compartment proportions
462 in SSc and esophageal motility phenotypes. It appeared that the low proportion of superficial cells in the
463 esophageal epithelium was a universal feature in SSc, regardless of phenotype (**Supplemental Figure 7,**
464 **A-B**). The proportion of superficial cells decreased with disease duration, but the trend was not statistically
465 significant (P=0.1; **Supplemental Figure 7**C). We also did not observe any significant correlations between
466 EEC compartment proportions and quantitative trait PCs (**Figure 5C**).

467 We then tested for correlations with gene expression in superficial EECs for genes that were
468 significantly differentially expressed in SSc compared to HCs (P_{adj}<0.05) and nominally differentially
469 expressed in SSc compared to GERD (P<0.05). There were 433 genes that met these criteria in the proximal
470 esophagus and 99 in the distal esophagus. Although no associations with esophageal motility phenotypes
471 were statistically significant after adjusting for multiple testing, the strongest phenotypic correlation was
472 with *TRIM11* (P_{FLIP}=0.01, P_{HRM}=0.007, **Supplemental Figure 7D**), a gene recently found to attenuate Treg
473 cell differentiation in $CD4^+$ T cells in mice[81]. No associations with the top two clinical PCs were significant
474 after adjusting for multiple testing. We also assessed clinical PC correlations with aggregate
475 metallothionein gene expression and transcription factor target expression for IRF1, MYC, E2F4, and
476 NFE2L2. We observed one nominal association between the mean metallothionein module score and PC1
477 (P=0.02, **Figure 5D**). Notably, within superficial cells we also observed a strong correlation between the
478 proportion of metallothionein-expressing cluster 3 cells and PC1 in the proximal (P=0.027) and distal
479 esophagus ($P=9.7\times10^{-6}$; **Figure 5E**). This indicates that not only is this population of cells reduced in SSc,
480 but its relative decrease is further correlated with more severe esophageal involvement (**Figure 5F**).
481
482

483 **DISCUSSION**

484 Esophageal dysfunction is extremely common in SSc and is significantly associated with increased
485 mortality and lower quality of life[82,83]. The causal mechanisms by which SSc affects the esophagus,
486 however, have not been conclusively determined. Here, we sought to clarify the role of epithelial cells in
487 SSc esophageal dysfunction by using scRNA-seq to quantify cellular and transcriptional changes relative
488 to HCs and individuals with GERD. While our findings suggest that epithelial changes in SSc result
489 primarily from chronic acid exposure, they also highlight immunoregulatory pathways uniquely altered in
490 SSc that may be linked to pathogenic aberrations.

491 The pathogenic role of epithelial cells in SSc has been an ongoing area of research[23]. When
492 damaged, epithelial cells release signals that help induce fibroblast activation to promote wound healing[84].
493 In SSc, epidermal keratinocyte characteristics resemble a pro-fibrotic, activated state[85]. SSc epidermal cells
494 were also found to stimulate fibroblasts in culture[86], and some signs related to epithelial-mesenchymal
495 transition (EMT) have been observed in the SSc epidermis[87]. The FLI1 (friend leukemia integration 1)
496 transcription factor, in particular, has received attention for its potential role in epithelial-cell-mediated SSc
497 pathogenesis[23]. Lower FLI1 expression was observed in the epidermis of diffuse cutaneous SSc, and
498 inactivation of FLI1 in human keratinocytes in vitro induced gene expression changes characteristic of
499 SSc[25]. Furthermore, conditional deletion of *Fli1* in epithelial cells produced an SSc-like phenotype in
500 mice[25]. Importantly, the epithelial-cell-specific *Fli1* knockout further recapitulated the esophageal
501 involvement of SSc, with increased collagen deposition in the lamina propria combined with atrophy of the
502 circular muscle layer, and the esophageal changes were not mediated by T-cell autoimmunity[25]. In contrast
503 to these previous studies in mice and epidermal keratinocytes, we did not observe *FLI1* expression in EECs
504 in any condition. *FLI1* was expressed at low levels in endothelial cells, but the expression differences
505 between SSc, HCs, and GERD were not significant. We did observe significantly increased expression in
506 SSc and GERD of the *PI3* gene, which encodes trappin-2/elafin, previously found to be induced by Fli1
507 silencing in human dermal microvascular endothelial cells[88]. However, this association is unlikely to be
508 pathogenic, as the increase in gene expression was greater in GERD, and a growing body of evidence

509 indicates that trappin-2/elafin is expressed to promote tissue repair in response to gastrointestinal tissue
510 inflammation[89].

511 While there have been many studies of epithelial cells in skin in SSc, molecular study of the human
512 esophagus in SSc has heretofore been limited to one array-based gene expression study in bulk tissue
513 conducted by Taroni and colleagues[15], in which they identified distinct expression signatures among
514 variable genes between 15 individuals with SSc that were similar to those seen in skin[90]. These signatures
515 were independent of clinically defined SSc subsets[15], suggestive of inter-individual heterogeneity in SSc.
516 They also found that gene expression patterns between upper and lower esophageal biopsies within
517 individuals were nearly identical. They did not observe statistical associations between inflammatory gene
518 expression signatures and GERD, but the study was underpowered for such an analysis, and GERD was
519 defined based on histological evidence for basal cell hyperplasia and intraepithelial lymphocyte counts.

520 The gene expression differences we observed in SSc epithelial cells were highly correlated with
521 those seen in GERD, and nearly all differential gene expression was limited to the superficial layers of the
522 epithelium. These observations suggest that the primary driver of differential gene expression in EECs in
523 SSc is chronic acid exposure. The genes with the strongest mutual upregulation in SSc and GERD were
524 primarily related to keratinization/cornification, including small proline-rich proteins (SPRRs), serine
525 protease inhibitors (serpins), S100 proteins, late cornified envelope (LCE) genes, and keratins associated
526 with terminal epithelial differentiation, namely *KRTDAP* and *KRT*. *KRT1* helps maintain epithelial barrier
527 function in gastrointestinal epithelial cells, and its overexpression was shown to attenuate IL-1β-induced
528 epithelial permeability[91]. The mutual upregulation of these genes in the outer layers of the epithelium
529 therefore likely represent a protective response to chronic acid exposure.

530 While the overall gene expression differences were highly correlated between SSc and GERD,
531 there were sets of genes disproportionately upregulated in SSc that hint at disease-related immune response
532 aberrations, including genes that have previously been implicated in SSc pathogenesis. For example,
533 *PTGES*, which encodes prostaglandin E synthase, was previously found to be significantly upregulated in
534 SSc fibroblasts[92] and inflammatory non-classical monocytes[93], and Ptges-null mice were resistant to

535 bleomycin-induced fibrosis[92]. CD44 and CD74 form the receptor complex for the macrophage migration
536 inhibitory factor (MIF), an inflammatory cytokine that promotes fibroblast migration and has been
537 implicated in SSc pathogenesis[94]. LTB4R (Leukotriene B4 Receptor, or BLT1) has been found to activate
538 AKT/mTOR signaling and knockdown of *LTB4R* attenuated fibrosis in bleomycin-induced SSc in murine
539 models[95]. Serum HBEGF (heparin-binding epidermal growth factor) levels and fibroblast *HBEGF*
540 expression were both significantly elevated in SSc[96]. Copy number variation of *APOBEC3A* was
541 significantly associated with SSc in a Han Chinese population[97]. Finally, the SSc-specific upregulation of
542 genes involved in antigen processing and presentation (*HLA-B*, *CD74*, *TAP1*, *PSMB8*, *PSMB9*) are
543 suggestive of increased human leukocyte antigen (HLA) class I activity in superficial EECs in SSc. Notably,
544 *TAP1*, *PSMB8*, and *PSMB9* are located immediately adjacent to one another in the class II region of the
545 HLA locus, which is the genomic region with the strongest genetic associations with SSc[98], and SSc-
546 associated HLA-B alleles have also been identified[99,100].

547 The most striking set of genes that were uniquely downregulated in SSc were metallothioneins.
548 The expression of detected metallothioneins (*MT1A*, *MT1E*, *MT1F*, *MT1G*, *MT1H*, *MT1M*, *MT1X*, *MT2A*)
549 was strongly reduced in the superficial compartment of EECs in the proximal esophagus in SSc compared
550 to HCs. We further observed a nominal association between relative metallothionein expression in the
551 proximal esophagus and the first PC of esophageal dysmotility among SSc patients, with lower expression
552 of metallothioneins correlated with greater EGJ distensibility and weaker contractility. The proportion of
553 the primary metallothionein-expressing cluster within superficial EECs was likewise significantly
554 associated with clinical PC1. The metallothioneins are a family of well-conserved, metal-binding proteins
555 that regulate zinc and copper homeostasis, prevent heavy metal poisoning, and combat oxidative
556 stress[101,102]. Their study in a wide array of physiological conditions and immune-mediated diseases has
557 revealed the proteins play central roles in innate and adaptive immunity, including autoimmunity, but their
558 immunoregulatory behavior is highly complex and context-specific[101,102]. Given their import in immune
559 regulation, their associations with SSc and esophageal motility reported in this study, and the connection

560 between SSc and heavy metal exposure[103], the metallothioneins are compelling candidates that warrant
561 further study in SSc.

562 Interestingly, there was greater relative gene dysregulation in the proximal esophagus compared to
563 the distal esophagus in SSc, but the opposite was true for GERD. This discrepancy suggests that esophageal
564 dysmotility in SSc leads to greater acid exposure in the proximal esophagus, since refluxate typically affects
565 the distal esophagus significantly more than the proximal esophagus[104] and abnormal peristalsis prolongs
566 acid clearance[105]. This finding coincides with a bulk RNA-seq study of esophageal mucosa of achalasia
567 patients[106], which likewise identified more differential expression in the proximal esophagus than in the
568 distal esophagus, including significant enrichment of matrisome-associated genes and lower expression of
569 genes associated with reactive oxygen species metabolism[106]. Due to differences in innervation depth[107],
570 the epithelium of the proximal esophagus has more nociceptive sensitivity than the distal esophagus, which
571 likely contributes to the association between gastrointestinal symptoms and lower health-related quality of
572 life in SSc[108]. Proximal acid exposure increases the risk of aspiration[109] and is significantly more prevalent
573 in SSc patients with idiopathic lung fibrosis[110].

574 Targets for several transcription factors, including IRF1, MYC, E2F4, and NFE2L2 (or NRF2),
575 were significantly enriched among DEGs in superficial EECs from the proximal esophagus. IRF1 is
576 involved in innate immune responses, but while IRF1 targets were enriched in DEGs between SSc and
577 GERD, the enrichment appeared to be driven by upregulation in GERD. IRF1 targets were not enriched in
578 DEGs between SSc and HCs, nor was *IRF1* significantly differentially expressed in SSc. MYC, E2F4, and
579 NFE2L2 all participate in different stages of the cell cycle[111-113], and expression of each was highest in
580 either the proliferating basal or suprabasal compartments, but these proteins may also play a role in terminal
581 differentiation of EECs, particularly under conditions of stress. In the least differentiated superficial cluster
582 (cluster 5), where the expression differences relative to HCs for each of these transcription factors and their
583 targets was most pronounced, the expression level changes in GERD were correlated with those observed
584 in SSc. Furthermore, Myc ablation in Krt14-expressing tissues in mice led to abnormal terminal
585 differentiation of epithelial cells[114], and studies of the esophageal epithelium in Nrf2-knockout mice

586 indicate that Nrf2 regulates cornification of keratinized epithelial cells under oxidative stress[115,116].
587 Moreover, *C10orf99*—one of the genes uniquely upregulated in SSc with superficial expression mirroring
588 *MYC*, *E2F4*, and *NFE2L2*—was found to increase expression of proinflammatory markers and attenuate
589 late differentiation of keratinocytes under conditions of stress[117]. These findings therefore suggest that
590 within the SSc esophagus there is activation of pathways that regulate the late differentiation of EECs in
591 response to external stress, and these could contribute to the reduction of superficial EECs we observed in
592 SSc.

593 There are important limitations to consider when interpreting this study. First, this study only
594 focused on squamous epithelial cells. There may be relevant changes in other cell types within the
595 esophageal mucosa that correlate with SSc and dysmotility, but such analyses are ongoing and were outside
596 the scope of this investigation. Second, the relatively small patient sample size of this study combined with
597 the heterogeneity of SSc subtypes and dysmotility phenotypes limited our power to detect clinical
598 associations between cell type proportions or gene expression changes and specific clinical phenotypes.
599 Heterogeneous molecular profiles in SSc esophagus samples have been described previously[15], but we
600 grouped all SSc patients together in our analyses for comparisons against GERD and HCs. We applied
601 ordinal logistic regression on motility classifications and performed principal components analysis on
602 reflux- and motility-related quantitative traits to increase our power to detect clinical associations but were
603 nonetheless statistically limited by sample size. Third, our analysis of imaging data in the SSc esophageal
604 mucosa was limited. While the spatial molecular imaging confirmed our cell annotations and reproduced
605 our findings within the superficial compartment, CosMx's cell segmentation algorithm could not aptly
606 capture the flattened, densely packed outer cell layers of the epithelium, resulting in a relatively sparse
607 representation of superficial EECs. Furthermore, the spatial analysis was limited to 8 samples from 4
608 individuals. Therefore, we cannot make more robust conclusions on the morphological states that
609 accompany the transcriptional and cellular proportion changes that we observed. Future studies utilizing
610 larger, more homogeneous cohorts with complete imaging data may be able to detect more esophageal
611 epithelial changes associated with dysmotility in SSc.

612 In summary, esophageal complications are common in individuals with SSc and can greatly impact
613 quality of life and lifespan. In this study, we sought to deepen our understanding of the role of the epithelium
614 in SSc esophageal involvement. Through a thorough, single-cell-level transcriptomic investigation, we
615 identified the unique cellular and transcriptional differences present in the SSc esophageal epithelium.
616 There were significantly fewer terminally differentiated epithelial cells in the apical, superficial layers in
617 both the proximal and distal esophagus, but otherwise epithelial compartment proportions were similar to
618 healthy controls, including proliferating cell proportions. Significant differences in gene expression were
619 likewise almost exclusively found in the superficial compartment, and these changes were highly correlated
620 with those seen in individuals with GERD, indicating that the primary driver of differential gene expression
621 in the SSc esophageal epithelium is chronic acid exposure. Transcriptional differences were more
622 prominent in the proximal esophagus of SSc patients than in GERD patients, possibly due to greater
623 proximal exposure resulting from esophageal dysmotility. Genes that were most disproportionately
624 dysregulated in SSc compared to GERD belonged to immune-related pathways, including innate and
625 adaptive response, antigen presentation, and metal homeostasis, pointing to pathogenic immune
626 dysfunction. Future studies investigating these pathways in non-epithelial cell populations in more
627 homogeneous SSc cohorts may further resolve the pathogenesis of esophageal involvement in SSc. By
628 serving as an atlas for the human esophageal epithelium in SSc, this study can guide future efforts to address
629 remaining gaps in our understanding of the SSc esophagus, to ultimately uncover pathogenic mechanisms
630 and identify actionable targets.

REFERENCES

1. Bairkdar, M. *et al.* Incidence and prevalence of systemic sclerosis globally: a comprehensive systematic review and meta-analysis. *Rheumatology (Oxford)* **60**, 3121-3133 (2021).
2. Fan, Y., Bender, S., Shi, W. & Zoz, D. Incidence and prevalence of systemic sclerosis and systemic sclerosis with interstitial lung disease in the United States. *J Manag Care Spec Pharm* **26**, 1539-1547 (2020).
3. Tian, J. *et al.* Global, regional, and national incidence and prevalence of systemic sclerosis. *Clin Immunol* **248**, 109267 (2023).
4. Pope, J.E. *et al.* Systemic Sclerosis and Associated Interstitial Lung Disease in Ontario, Canada: An Examination of Prevalence and Survival Over 10 Years. *J Rheumatol* **48**, 1427-1434 (2021).
5. Li, J.X. Secular Trends in Systemic Sclerosis Mortality in the United States from 1981 to 2020. *Int J Environ Res Public Health* **19**(2022).
6. Rubio-Rivas, M., Royo, C., Simeon, C.P., Corbella, X. & Fonollosa, V. Mortality and survival in systemic sclerosis: systematic review and meta-analysis. *Semin Arthritis Rheum* **44**, 208-19 (2014).
7. Bairkdar, M. *et al.* Survival in Swedish patients with systemic sclerosis: a nationwide population-based matched cohort study. *Rheumatology (Oxford)* **62**, 1170-1178 (2023).
8. Li, B. *et al.* Esophageal Dysfunction in Systemic Sclerosis: An Update. *Rheumatol Ther* **8**, 1535-1549 (2021).
9. Emmanuel, A. Current management of the gastrointestinal complications of systemic sclerosis. *Nat Rev Gastroenterol Hepatol* **13**, 461-72 (2016).
10. Ahuja, N.K. & Clarke, J.O. Scleroderma and the Esophagus. *Gastroenterol Clin North Am* **50**, 905-918 (2021).
11. Denaxas, K., Ladas, S.D. & Karamanolis, G.P. Evaluation and management of esophageal manifestations in systemic sclerosis. *Ann Gastroenterol* **31**, 165-170 (2018).
12. Foocharoen, C. *et al.* Development and validation of machine learning for early mortality in systemic sclerosis. *Sci Rep* **12**, 17178 (2022).
13. Sjogren, R.W. Gastrointestinal motility disorders in scleroderma. *Arthritis Rheum* **37**, 1265-82 (1994).
14. Roberts, C.G., Hummers, L.K., Ravich, W.J., Wigley, F.M. & Hutchins, G.M. A case-control study of the pathology of oesophageal disease in systemic sclerosis (scleroderma). *Gut* **55**, 1697-703 (2006).
15. Taroni, J.N. *et al.* Molecular characterization of systemic sclerosis esophageal pathology identifies inflammatory and proliferative signatures. *Arthritis Res Ther* **17**, 194 (2015).
16. Roman, S. *et al.* Esophageal dysmotility associated with systemic sclerosis: a high-resolution manometry study. *Dis Esophagus* **24**, 299-304 (2011).
17. Noviani, M., Chellamuthu, V.R., Albani, S. & Low, A.H.L. Toward Molecular Stratification and Precision Medicine in Systemic Sclerosis. *Front Med (Lausanne)* **9**, 911977 (2022).
18. Carlson, D.A. *et al.* Loss of Peristaltic Reserve, Determined by Multiple Rapid Swallows, Is the Most Frequent Esophageal Motility Abnormality in Patients With Systemic Sclerosis. *Clin Gastroenterol Hepatol* **14**, 1502-6 (2016).
19. Kimmel, J.N. *et al.* The association between systemic sclerosis disease manifestations and esophageal high-resolution manometry parameters. *Neurogastroenterol Motil* **28**, 1157-65 (2016).
20. Carlson, D.A. *et al.* Heterogeneity of primary and secondary peristalsis in systemic sclerosis: A new model of "scleroderma esophagus". *Neurogastroenterol Motil* **34**, e14284 (2022).
21. Durcan, C. *et al.* Experimental investigations of the human oesophagus: anisotropic properties of the embalmed mucosa-submucosa layer under large deformation. *Biomech Model Mechanobiol* **21**, 1685-1702 (2022).
22. Kou, W., Pandolfino, J.E., Kahrilas, P.J. & Patankar, N.A. Simulation studies of the role of esophageal mucosa in bolus transport. *Biomech Model Mechanobiol* **16**, 1001-1009 (2017).

680 23. Asano, Y., Takahashi, T. & Saigusa, R. Systemic sclerosis: Is the epithelium a missing piece of the
681 pathogenic puzzle? *J Dermatol Sci* **94**, 259-265 (2019).
682 24. Asano, Y., Bujor, A.M. & Trojanowska, M. The impact of Fli1 deficiency on the pathogenesis of
683 systemic sclerosis. *J Dermatol Sci* **59**, 153-62 (2010).
684 25. Takahashi, T. *et al.* Epithelial Fli1 deficiency drives systemic autoimmunity and fibrosis: Possible
685 roles in scleroderma. *J Exp Med* **214**, 1129-1151 (2017).
686 26. Dinh, H.Q. *et al.* Integrated single-cell transcriptome analysis reveals heterogeneity of esophageal
687 squamous cell carcinoma microenvironment. *Nat Commun* **12**, 7335 (2021).
688 27. Zhang, X. *et al.* Dissecting esophageal squamous-cell carcinoma ecosystem by single-cell
689 transcriptomic analysis. *Nat Commun* **12**, 5291 (2021).
690 28. Rochman, M. *et al.* Single-cell RNA-Seq of human esophageal epithelium in homeostasis and
691 allergic inflammation. *JCI Insight* **7**(2022).
692 29. Clevenger, M.H. *et al.* Suprabasal cells retain progenitor cell identity programs in eosinophilic
693 esophagitis-driven basal cell hyperplasia. *JCI Insight* **8**(2023).
694 30. Carlson, D.A. *et al.* Classifying Esophageal Motility by FLIP Panometry: A Study of 722 Subjects
695 With Manometry. *Am J Gastroenterol* **116**, 2357-2366 (2021).
696 31. Koop, A.H., Kahrilas, P.J., Schauer, J., Pandolfino, J.E. & Carlson, D.A. The impact of primary
697 peristalsis, contractile reserve, and secondary peristalsis on esophageal clearance measured by
698 timed barium esophagogram. *Neurogastroenterol Motil*, e14638 (2023).
699 32. Stern, E.K. *et al.* Abnormal esophageal acid exposure on high-dose proton pump inhibitor therapy
700 is common in systemic sclerosis patients. *Neurogastroenterol Motil* **30**(2018).
701 33. Yadlapati, R. *et al.* Esophageal motility disorders on high-resolution manometry: Chicago
702 classification version 4.0((c)). *Neurogastroenterol Motil* **33**, e14058 (2021).
703 34. Rooney, K.P. *et al.* Esophagogastric Junction Opening Parameters Are Consistently Abnormal in
704 Untreated Achalasia. *Clin Gastroenterol Hepatol* **19**, 1058-1060.e1 (2021).
705 35. Carlson, D.A. *et al.* Evaluating esophageal motility beyond primary peristalsis: Assessing
706 esophagogastric junction opening mechanics and secondary peristalsis in patients with normal
707 manometry. *Neurogastroenterol Motil* **33**, e14116 (2021).
708 36. Carlson, D.A. *et al.* Normal Values of Esophageal Distensibility and Distension-Induced
709 Contractility Measured by Functional Luminal Imaging Probe Panometry. *Clin Gastroenterol*
710 *Hepatol* **17**, 674-681 e1 (2019).
711 37. Jonasson, C., Wernersson, B., Hoff, D.A. & Hatlebakk, J.G. Validation of the GerdQ questionnaire
712 for the diagnosis of gastro-oesophageal reflux disease. *Aliment Pharmacol Ther* **37**, 564-72 (2013).
713 38. Taft, T.H. *et al.* Development and validation of the brief esophageal dysphagia questionnaire.
714 *Neurogastroenterol Motil* **28**, 1854-1860 (2016).
715 39. Bedell, A., Taft, T.H., Keefer, L. & Pandolfino, J. Development of the Northwestern Esophageal
716 Quality of Life Scale: A Hybrid Measure for Use Across Esophageal Conditions. *Am J*
717 *Gastroenterol* **111**, 493-9 (2016).
718 40. Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587 e29 (2021).
719 41. Hippen, A.A. *et al.* miQC: An adaptive probabilistic framework for quality control of single-cell
720 RNA-sequencing data. *PLoS Comput Biol* **17**, e1009290 (2021).
721 42. Germain, P.L., Lun, A., Garcia Meixide, C., Macnair, W. & Robinson, M.D. Doublet identification
722 in single-cell sequencing data using scDblFinder. *F1000Res* **10**, 979 (2021).
723 43. Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-seq data
724 using regularized negative binomial regression. *Genome Biol* **20**, 296 (2019).
725 44. Lause, J., Berens, P. & Kobak, D. Analytic Pearson residuals for normalization of single-cell RNA-
726 seq UMI data. *Genome Biol* **22**, 258 (2021).
727 45. McInnes, L., Healy, J. & Melville, J. UMAP: Uniform Manifold Approximation and Projection for
728 Dimension Reduction. arXiv:1802.03426 (2018).
729 46. Busslinger, G.A. *et al.* Human gastrointestinal epithelia of the esophagus, stomach, and duodenum
730 resolved at single-cell resolution. *Cell Rep* **34**, 108819 (2021).

731 47. Kabir, M.F. *et al.* Single cell transcriptomic analysis reveals cellular diversity of murine esophageal
732 epithelium. *Nat Commun* **13**, 2167 (2022).
733 48. Okumura, T., Shimada, Y., Imamura, M. & Yasumoto, S. Neurotrophin receptor p75(NTR)
734 characterizes human esophageal keratinocyte stem cells in vitro. *Oncogene* **22**, 4017-26 (2003).
735 49. Tirosh, I. *et al.* Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-
736 seq. *Science* **352**, 189-96 (2016).
737 50. Fonseka, C.Y. *et al.* Mixed-effects association of single cells identifies an expanded effector
738 CD4(+) T cell subset in rheumatoid arthritis. *Sci Transl Med* **10**(2018).
739 51. Mootha, V.K. *et al.* PGC-1alpha-responsive genes involved in oxidative phosphorylation are
740 coordinately downregulated in human diabetes. *Nat Genet* **34**, 267-73 (2003).
741 52. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting
742 genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545-50 (2005).
743 53. Liberzon, A. *et al.* The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell*
744 *Syst* **1**, 417-425 (2015).
745 54. Chen, Y., Lun, A.T. & Smyth, G.K. From reads to genes to pathways: differential expression
746 analysis of RNA-Seq experiments using Rsubread and the edgeR quasi-likelihood pipeline.
747 *F1000Res* **5**, 1438 (2016).
748 55. McCarthy, D.J., Chen, Y. & Smyth, G.K. Differential expression analysis of multifactor RNA-Seq
749 experiments with respect to biological variation. *Nucleic Acids Res* **40**, 4288-97 (2012).
750 56. Robinson, M.D., McCarthy, D.J. & Smyth, G.K. edgeR: a Bioconductor package for differential
751 expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-40 (2010).
752 57. Wu, T. *et al.* clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation*
753 *(Camb)* **2**, 100141 (2021).
754 58. Consortium, E.P. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**,
755 57-74 (2012).
756 59. Lachmann, A. *et al.* ChEA: transcription factor regulation inferred from integrating genome-wide
757 ChIP-X experiments. *Bioinformatics* **26**, 2438-44 (2010).
758 60. Kou, Y. *et al.* ChEA2: Gene-Set Libraries from ChIP-X Experiments to Decode the Transcription
759 Regulome. 416-430 (Springer Berlin Heidelberg, Berlin, Heidelberg, 2013).
760 61. Chen, E.Y. *et al.* Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool.
761 *BMC Bioinformatics* **14**, 128 (2013).
762 62. Aran, D. *et al.* Reference-based analysis of lung single-cell sequencing reveals a transitional
763 profibrotic macrophage. *Nat Immunol* **20**, 163-172 (2019).
764 63. Koop, A.H., Kahrilas, P.J., Schauer, J., Pandolfino, J.E. & Carlson, D.A. The impact of primary
765 peristalsis, contractile reserve, and secondary peristalsis on esophageal clearance measured by
766 timed barium esophagogram. *Neurogastroenterol Motil* **35**, e14638 (2023).
767 64. Squair, J.W. *et al.* Confronting false discoveries in single-cell differential expression. *Nat Commun*
768 **12**, 5692 (2021).
769 65. Wang, Q. *et al.* Targeting microsomal prostaglandin E synthase 1 to develop drugs treating the
770 inflammatory diseases. *Am J Transl Res* **13**, 391-419 (2021).
771 66. Yi, Y.S. Functional Role of Milk Fat Globule-Epidermal Growth Factor VIII in Macrophage-
772 Mediated Inflammatory Responses and Inflammatory/Autoimmune Diseases. *Mediators Inflamm*
773 **2016**, 5628486 (2016).
774 67. Baaten, B.J., Li, C.R. & Bradley, L.M. Multifaceted regulation of T cells by CD44. *Commun Integr*
775 *Biol* **3**, 508-12 (2010).
776 68. Liu, Q. *et al.* Role of the mucin-like glycoprotein FCGBP in mucosal immunity and cancer. *Front*
777 *Immunol* **13**, 863317 (2022).
778 69. Oh, S. *et al.* Genotoxic stress and viral infection induce transient expression of APOBEC3A and
779 pro-inflammatory genes through two distinct pathways. *Nat Commun* **12**, 4917 (2021).
780 70. Tiwari, R., de la Torre, J.C., McGavern, D.B. & Nayak, D. Beyond Tethering the Viral Particles:
781 Immunomodulatory Functions of Tetherin (BST-2). *DNA Cell Biol* **38**, 1170-1177 (2019).

782 71. Nguyen, H.T. *et al.* CXCR7: a beta-arrestin-biased receptor that potentiates cell migration and
783 recruits beta-arrestin2 exclusively through Gbetagamma subunits and GRK2. *Cell Biosci* **10**, 134
784 (2020).
785 72. Okamoto, Y. & Shikano, S. Emerging roles of a chemoattractant receptor GPR15 and ligands in
786 pathophysiology. *Front Immunol* **14**, 1179456 (2023).
787 73. Subramanian, B.C., Moissoglu, K. & Parent, C.A. The LTB(4)-BLT1 axis regulates the polarized
788 trafficking of chemoattractant GPCRs during neutrophil chemotaxis. *J Cell Sci* **131**(2018).
789 74. Kelly, A. & Trowsdale, J. Genetics of antigen processing and presentation. *Immunogenetics* **71**,
790 161-170 (2019).
791 75. Bartel, Y., Bauer, B. & Steinle, A. Modulation of NK cell function by genetically coupled C-type
792 lectin-like receptor/ligand pairs encoded in the human natural killer gene complex. *Front Immunol*
793 **4**, 362 (2013).
794 76. Li, Y. *et al.* HB-EGF-induced IL-8 secretion from airway epithelium leads to lung fibroblast
795 proliferation and migration. *BMC Pulm Med* **21**, 347 (2021).
796 77. Lu, R.Q. *et al.* SGK1, a Critical Regulator of Immune Modulation and Fibrosis and a Potential
797 Therapeutic Target in Chronic Graft-Versus-Host Disease. *Front Immunol* **13**, 822303 (2022).
798 78. Uhlen, M. *et al.* Proteomics. Tissue-based map of the human proteome. *Science* **347**, 1260419
799 (2015).
800 79. Politi, E. *et al.* Filaggrin and Periostin Expression Is Altered in Eosinophilic Esophagitis and
801 Normalized With Treatment. *J Pediatr Gastroenterol Nutr* **65**, 47-52 (2017).
802 80. Oshima, N. *et al.* Epidermal differentiation complex protein involucrin is down-regulated in
803 eosinophilic esophagitis. *Esophagus* **14**, 171-177 (2017).
804 81. Yu, T. *et al.* TRIM11 attenuates Treg cell differentiation by p62-selective autophagic degradation
805 of AIM2. *Cell Rep* **42**, 113231 (2023).
806 82. Crowell, M.D. *et al.* Esophageal Motor Abnormalities in Patients With Scleroderma:
807 Heterogeneity, Risk Factors, and Effects on Quality of Life. *Clin Gastroenterol Hepatol* **15**, 207-
808 213 e1 (2017).
809 83. Forbes, A. & Marie, I. Gastrointestinal complications: the most frequent internal complications of
810 systemic sclerosis. *Rheumatology (Oxford)* **48 Suppl 3**, iii36-9 (2009).
811 84. Krieg, T., Abraham, D. & Lafyatis, R. Fibrosis in connective tissue disease: the role of the
812 myofibroblast and fibroblast-epithelial cell interactions. *Arthritis Res Ther* **9 Suppl 2**, S4 (2007).
813 85. Aden, N. *et al.* Proteomic analysis of scleroderma lesional skin reveals activated wound healing
814 phenotype of epidermal cell layer. *Rheumatology (Oxford)* **47**, 1754-60 (2008).
815 86. Aden, N. *et al.* Epithelial cells promote fibroblast activation via IL-1alpha in systemic sclerosis. *J*
816 *Invest Dermatol* **130**, 2191-200 (2010).
817 87. Nikitorowicz-Buniak, J., Denton, C.P., Abraham, D. & Stratton, R. Partially Evoked Epithelial-
818 Mesenchymal Transition (EMT) Is Associated with Increased TGFbeta Signaling within Lesional
819 Scleroderma Skin. *PLoS One* **10**, e0134092 (2015).
820 88. Miyagawa, T. *et al.* A potential contribution of trappin-2 to the development of vasculopathy in
821 systemic sclerosis. *J Eur Acad Dermatol Venereol* **33**, 753-760 (2019).
822 89. Deraison, C., Bonnart, C., Langella, P., Roget, K. & Vergnolle, N. Elafin and its precursor trappin-
823 2: What is their therapeutic potential for intestinal diseases? *Br J Pharmacol* **180**, 144-160 (2023).
824 90. Hinchcliff, M. *et al.* Molecular signatures in skin associated with clinical improvement during
825 mycophenolate treatment in systemic sclerosis. *J Invest Dermatol* **133**, 1979-89 (2013).
826 91. Dong, X. *et al.* Critical role of Keratin 1 in maintaining epithelial barrier and correlation of its
827 down-regulation with the progression of inflammatory bowel disease. *Gene* **608**, 13-19 (2017).
828 92. McCann, M.R. *et al.* mPGES-1 null mice are resistant to bleomycin-induced skin fibrosis. *Arthritis*
829 *Res Ther* **13**, R6 (2011).
830 93. Villanueva-Martin, G. *et al.* Non-classical circulating monocytes expressing high levels of
831 microsomal prostaglandin E2 synthase-1 tag an aberrant IFN-response in systemic sclerosis. *J*
832 *Autoimmun* **140**, 103097 (2023).

833 94. Unlu, B., Tursen, U., Rajabi, Z., Jabalameli, N. & Rajabi, F. The Immunogenetics of Systemic
834 Sclerosis. *Adv Exp Med Biol* **1367**, 259-298 (2022).
835 95. Liang, M. *et al.* Promotion of Myofibroblast Differentiation and Tissue Fibrosis by the Leukotriene
836 B(4) -Leukotriene B(4) Receptor Axis in Systemic Sclerosis. *Arthritis Rheumatol* **72**, 1013-1025
837 (2020).
838 96. Hirabayashi, M. *et al.* Possible pro-inflammatory role of heparin-binding epidermal growth factor-
839 like growth factor in the active phase of systemic sclerosis. *J Dermatol* **45**, 182-188 (2018).
840 97. Guo, S. *et al.* Copy Number Variation of HLA-DQA1 and APOBEC3A/3B Contribute to the
841 Susceptibility of Systemic Sclerosis in the Chinese Han Population. *J Rheumatol* **43**, 880-6 (2016).
842 98. Lopez-Isac, E. *et al.* GWAS for systemic sclerosis identifies multiple risk loci and highlights
843 fibrotic and vasculopathy pathways. *Nat Commun* **10**, 4955 (2019).
844 99. Acosta-Herrera, M. *et al.* Comprehensive analysis of the major histocompatibility complex in
845 systemic sclerosis identifies differential HLA associations by clinical and serological subtypes. *Ann*
846 *Rheum Dis* **80**, 1040-1047 (2021).
847 100. Hanson, A.L. *et al.* Contribution of HLA and KIR Alleles to Systemic Sclerosis Susceptibility and
848 Immunological and Clinical Disease Subtypes. *Front Genet* **13**, 913196 (2022).
849 101. Dai, H., Wang, L., Li, L., Huang, Z. & Ye, L. Metallothionein 1: A New Spotlight on Inflammatory
850 Diseases. *Front Immunol* **12**, 739918 (2021).
851 102. Subramanian Vignesh, K. & Deepe, G.S., Jr. Metallothioneins: Emerging Modulators in Immunity
852 and Infection. *Int J Mol Sci* **18**(2017).
853 103. Marie, I. Systemic sclerosis and exposure to heavy metals. *Autoimmun Rev* **18**, 62-72 (2019).
854 104. Shaker, R., Dodds, W.J., Helm, J.F., Kern, M.K. & Hogan, W.J. Regional esophageal distribution
855 and clearance of refluxed gastric acid. *Gastroenterology* **101**, 355-9 (1991).
856 105. Diener, U., Patti, M.G., Molena, D., Fisichella, P.M. & Way, L.W. Esophageal dysmotility and
857 gastroesophageal reflux disease. *J Gastrointest Surg* **5**, 260-5 (2001).
858 106. Patel, C.K. *et al.* RNA-sequencing reveals molecular and regional differences in the esophageal
859 mucosa of achalasia patients. *Sci Rep* **12**, 20616 (2022).
860 107. Woodland, P. *et al.* Distinct afferent innervation patterns within the human proximal and distal
861 esophageal mucosa. *Am J Physiol Gastrointest Liver Physiol* **308**, G525-31 (2015).
862 108. Sandqvist, G. & Hesselstrand, R. Validity of the Swedish version of the systemic sclerosis quality
863 of life questionnaire (SSCQoL): A novel measure of quality of life for patients with systemic
864 sclerosis. *Ann Rheum Dis* **78**, 855-857 (2019).
865 109. Lee, A.S., Lee, J.S., He, Z. & Ryu, J.H. Reflux-Aspiration in Chronic Lung Disease. *Ann Am*
866 *Thorac Soc* **17**, 155-164 (2020).
867 110. Savarino, E. *et al.* Gastroesophageal reflux and pulmonary fibrosis in scleroderma: a study using
868 pH-impedance monitoring. *Am J Respir Crit Care Med* **179**, 408-13 (2009).
869 111. Conboy, C.M. *et al.* Cell cycle genes are the evolutionarily conserved targets of the E2F4
870 transcription factor. *PLoS One* **2**, e1061 (2007).
871 112. Garcia-Gutierrez, L., Delgado, M.D. & Leon, J. MYC Oncogene Contributions to Release of Cell
872 Cycle Brakes. *Genes (Basel)* **10**(2019).
873 113. Lastra, D., Escoll, M. & Cuadrado, A. Transcription Factor NRF2 Participates in Cell Cycle
874 Progression at the Level of G1/S and Mitotic Checkpoints. *Antioxidants (Basel)* **11**(2022).
875 114. Portal, C., Wang, Z., Scott, D.K., Wolosin, J.M. & Iomini, C. The c-Myc Oncogene Maintains
876 Corneal Epithelial Architecture at Homeostasis, Modulates p63 Expression, and Enhances
877 Proliferation During Tissue Repair. *Invest Ophthalmol Vis Sci* **63**, 3 (2022).
878 115. Wruck, C.J. *et al.* Impact of Nrf2 on esophagus epithelium cornification. *Int J Dermatol* **50**, 1362-
879 1365 (2011).
880 116. Chen, H. *et al.* Nrf2 deficiency impairs the barrier function of mouse oesophageal epithelium. *Gut*
881 **63**, 711-9 (2014).
882 117. Dainichi, T. *et al.* C10orf99/GPR15L Regulates Proinflammatory Response of Keratinocytes and
883 Barrier Formation of the Skin. *Front Immunol* **13**, 825032 (2022).

884 118. Jones, R. *et al.* Development of the GerdQ, a tool for the diagnosis and management of gastro-
885 oesophageal reflux disease in primary care. *Aliment Pharmacol Ther* **30**, 1030-8 (2009).
886

887 **FIGURES**

888 **Figure 1. ScRNA-seq sample composition by cell type. A)** Integrated UMAP embedding of all cells
889 (n=306,372) labeled by cell type. **B)** Proportion of cells by cell type. **C)** Expression by cell type of canonical
890 markers. **D)** Sample-wise (n=39) distribution of major cell types. **E)** Proportion of major non-epithelial cell
891 types by condition and biopsy location. Box whiskers extend to range of values.

892

893 **Figure 2. Landscape of esophageal epithelial cells. A)** Histological summary of the human esophageal
894 epithelium, adopted from Clevenger et al.[29] Alternatively shaded cells within the same layer represent
895 proliferating cells. NOTE: The layers of replicating cells not attached to the basement membrane, labeled
896 here as proliferating suprabasal cells, are sometimes referred to as epibasal cells[29]. **B)** Integrated UMAP
897 embedding of all esophageal epithelial cells (EECs, n=230,720) labeled by epithelial compartment. **C)**
898 Expression of EEC compartment and cell cycle markers by condition. **D)** CosMx spatial molecular imaging
899 of proximal and distal esophageal mucosal tissue in scleroderma with absent contractility, annotated using
900 label transfer from the scRNA-seq data. **E)** Proportions of cells by condition, EEC compartment, and biopsy
901 location. The bars denote the mean values, the vertical lines the standard deviations, and the points the
902 individual sample proportions. Pairwise differences across conditions were evaluated statistically using
903 MASC[50]. *$p<0.05$.

904

905 **Figure 3. Differential gene expression between conditions in esophageal epithelial cells. A)** The
906 number of significantly differentially expressed genes at the single-cell level with $|\log_2 FC| > 0.1$ are shown
907 for SSC vs. HCs (purple) and GERD vs. HCs (orange), by epithelial compartment and biopsy location. Blue
908 bars account for genes that were differentially expressed in both SSc and GERD compared to HCs. **B)**
909 Gene set enrichment analysis results, showing all pathways enriched with $P<0.001$ (unadjusted) for SSc
910 vs. HCs (purple) and GERD vs. HCs (orange), split by non-proliferating epithelial compartment and biopsy
911 location. Points circled in red indicate statistical significance after adjusting for the number of tested
912 pathways ($P_{FDR}<0.05$). Pathways in bold denote pathways with significant enrichment in SSc. **C)** The
913 distribution of intra-sample EEC gene expression correlations for the 2,000 most variable genes between
914 proximal and distal biopsies are shown by epithelial compartment and condition. Box whiskers extend to

range of values. **D)** Pseudobulk gene expression differences in $\log_2$FC are shown for SSc vs. HCs (Y axis) against GERD vs. HCs (X axis) within the superficial compartment in the proximal and distal esophagus regions. Significant DEGs in both conditions are highlighted in blue, significant DEGs in SSC only are highlighted in purple, and significant DEGs in GERD only are highlighted in orange. The correlation between comparisons is shown in the upper left, and the slope and coefficient of determination for the modeled linear regression with intercept=0 is displayed in the lower right. Trendlines with 95% confidence intervals are plotted in pink, and the dashed grey lines denote where y=x. Encircled points highlight disease-specific DEGs that are referenced in the main text and plotted in **Supplemental Figure 4H**, which shows the distribution of expression aggregated by sample in the superficial compartment across conditions and biopsy locations.

Figure 4. Landscape of superficial esophageal epithelial cells. A) Integrated UMAP embedding of all superficial EECs (n=30,866) clustered by unique transcriptional signatures. **B)** Relative expression of cluster-specific gene signatures. **C)** Superficial module score by superficial cell cluster. **D)** CosMx spatial molecular imaging of proximal esophageal mucosal tissue from a healthy control donor and in scleroderma with absent contractility, annotated using label transfer from the scRNA-seq data. Superficial cluster proportions were derived from all imaged proximal tissue (n=2 SSc, 2 HCs). **E)** Proportion of superficial cells by condition and cluster from the proximal esophagus. The bars denote the mean values, the vertical lines the standard deviations, and the points the individual sample proportions. Pairwise differences across conditions were evaluated statistically using MASC[50]. *$p<0.05$; **$p<0.005$. **F)** Distribution of metallothionein module score in proximal, superficial cells by condition. The metallothionein module score included *MT1A*, *MT1E*, *MT1F*, *MT1G*, *MT1H*, *MT1M*, *MT1X*, and *MT2A*. **G)** Distribution of relative metallothionein expression by cluster and condition in superficial cells from proximal esophagus. All pairwise comparisons between SSc and HCs were statistically significant, except for cluster 5. **H)** CosMx spatial molecular imaging of proximal esophageal mucosal tissue from a healthy control donor and in scleroderma with absent contractility. Metallothionein module scores are plotted within the superficial EECs. **I)** Distributions of metallothionein module scores by cell cluster and condition from the CosMx spatial molecular imaging

942 data. Clusters withs significant pairwise differences (Wilcoxon) in module score levels (P_{FDR}<0.05) are
943 denoted with an asterisk.
944
945 **Figure 5. Correlations with clinical phenotypes. A)** Heatmap of quantitative clinical traits organized using
946 agglomerative, complete hierarchical clustering on Euclidean distances, with pairwise Spearman
947 correlations shown in each cell. **B)** SSc cases are plotted on the first two PCs of the quantitative clinical
948 traits, colored by esophageal motility phenotype. The relative magnitude and direction of trait correlations
949 with the PCs are shown with black arrows. **C)** Correlations between EEC compartments and clinical trait
950 PCs in SSc cases. Linear regression trendlines with 95% confidence intervals are shown for each
951 compartment against PC1 and PC2. **D)** Correlations between aggregate, relative metallothionein
952 expression and clinical trait PCs in superficial cells in the proximal esophagus. Trendlines with 95%
953 confidence intervals are shown with the unadjusted correlation P-value. The metallothionein module score
954 included *MT1A*, *MT1E*, *MT1F*, *MT1G*, *MT1H*, *MT1M*, *MT1X*, and *MT2A*. **E)** Correlations between superficial
955 clusters (SCs) and clinical trait PC1 in SSc cases. Linear regression trendlines with 95% confidence
956 intervals are shown for each SC. The displayed P-values correspond to the highlighted correlations
957 between SC3 and PC1. **F)** Boxplots showing the proportion of SC3 in superficial cells by esophageal motility
958 phenotype in the proximal and distal esophagus, colored by disease state (HCs, grey; GERD, orange; SSc,
959 purple). AET, acid exposure time; BEDQ, brief esophageal dysphagia questionnaire; BMI, body mass index;
960 DCI, distal contractile integral; DI, distensibility index; EGJ, esophagogastric junction; EGJOO,
961 esophagogastric junction outflow obstruction; FLIP, functional luminal imaging probe panometry; HCs,
962 healthy controls; HRM, high-resolution manometry; IEM, ineffective esophageal motility; IQR, inter-quartile
963 range; IRP, integrated relaxation pressure; NEQOL, Northwestern esophageal quality of life; NM,
964 neuromyogenic model. NR, not reported.

965 **SUPPLEMENTAL FIGURES**

966 **Supplemental Figure 1. Cell clustering and cell type annotation. A)** Cell type module scores for 10
967 detected cell types, colored by relative gene expression of the listed genes. **B)** Projected UMAP
968 embeddings calculated on the top 40 PCs with n=200 neighbors. Cells are colored by cluster, which are
969 numbered in descending order by total cell count. Clusters were determined using Seurat's modularity-
970 based clustering on a shared nearest neighbor (SNN) graph with k=25 and resolution=0.25 on the top 40
971 PCs. **C)** Isolation of cluster 13 for annotation of cell types with smaller proportions of cells and projected
972 UMAP embeddings for cluster 13, colored by subcluster. **D)** Cell type module scores for 7 detected cell
973 types in subcluster 13, colored by relative gene expression of the listed genes.

974

975 **Supplemental Figure 2. Epithelial cell quality control and annotation. A)** EEC UMAP projection,
976 colored by the number of unique molecular identifiers (UMIs) per cell. **B)** EEC UMAP projection, colored by
977 the number of detected genes per cell. **C)** EEC UMAP projection, colored by the number of proportion of
978 UMIs mapping to mitochondrial DNA. **D)** EEC UMAP projection, colored by the proportion of UMIs mapping
979 to ribosomal DNA. **E)** Distribution of number of UMIs detected per cell. EECs with <3500 UMIs/cell were
980 removed from further analysis. **F)** EEC UMAPs colored by gene expression (log(counts+1))) for genes
981 included in compartment-specific expression scores. **G)** EEC UMAPs colored by compartment expression
982 scores. **H)** Projected UMAP embeddings for post-quality-control EECs calculated on the top 35 PCs with
983 n=100 neighbors. Cells are colored by cluster, which are numbered in descending order by total cell count.
984 Clusters were determined using Seurat's modularity-based clustering on a SNN graph with k=50 and
985 resolution=0.25 on the top 35 PCs. **I)** Result of K-means clustering to distinguish suprabasal from basal
986 cells, based on expression of corresponding cluster markers (**F**) in clusters 3, 4, and 5, with K=2. **J)** Result
987 of K-means clustering to distinguish superficial from suprabasal cells, based on expression of
988 corresponding cluster markers (**F**) in clusters 6, 7, and 8, with K=2. **K)** Predicted classification of cell cycle
989 state in EECs determined using Seurat's CellCycleScoring() function.

990

991

992 **Supplemental Figure 3. Spatial molecular imaging of esophageal mucosa. A)** Esophageal mucosa
993 tissue slides analyzed by CosMx spatial molecular imaging are shown with annotations conferred from
994 scRNA-seq label transfer for two healthy controls (HCs) and two SSc patients with absent contractility **B)**
995 Spatial molecular imaging cell type counts are shown by individual and biopsy location. En, endothelial; Ep,
996 epithelial; F, fibroblast; L, lymphocytes; Mc, mast cells; My, myeloid; P, pericytes; SMCs, smooth muscle
997 cells; SMG, submucosal glands.
998
999 **Supplemental Figure 4. Group-wise gene expression correlation for variable genes. A)** UpSet plots
1000 showing differential gene expression at the single-cell level by epithelial compartment for SSc vs HCs
1001 (purple) and GERD vs HCs (orange) in the proximal and distal esophagus. **B)** Multidimensional scaling
1002 plots of aggregate gene expression samples colored by sample, epithelial compartment, condition, and
1003 biopsy location. **C-D)** Pseudobulk gene expression differences between SSc and HCs **(C)** and between
1004 SSc and GERD **(D)** are shown for each EEC compartment. Significant DEGs (FDR<0.05) are highlighted
1005 in purple and raspberry, respectively. The percentages of genes that were significantly differentially
1006 expressed in each compartment are plotted in blue in along the secondary Y axis. **E)** The number of
1007 significant DEGs are shown for GERD vs. HCs (orange) and for all permutations of n=4 samples for SSc
1008 vs. HCs (purple) in superficial cells for both proximal and distal regions. The point within the SSc distribution
1009 is the median value. **F-G)** The distributions of relative FC or slope *m* **(F)** and coefficient of determination r^2
1010 **(G)** are shown for all permutations of n=4 SSc samples, determined by modeling a linear regression with
1011 intercept=0 for the $\log_2$FC of SSc vs. HCs against the $\log_2$FC of GERD vs. HCs for all expressed genes in
1012 the superficial compartment for both proximal and distal regions. DEG, differentially expressed gene; FC,
1013 fold change. **H)** Boxplots of pseudobulk gene expression, aggregated by condition, biopsy location, and
1014 EEC compartment for disease-specific DEGs referenced in the main text. Pairwise differential expression
1015 was determined using edgeR's quasi-likelihood F-tests: *p<0.05; **p<0.05; ***p<0.0005.
1016
1017 **Supplemental Figure 5. Expression of cluster markers, enriched transcription factors, and their**
1018 **targets in superficial EECs in proximal esophagus. A)** UMAPs with expression for the top superficial
1019 cluster markers are shown along with the epithelial differentiation score for superficial cells. **B-D)** The

1020 distributions of enriched transcription factor expression and target module scores are shown proximal,
1021 superficial cells by condition and superficial cluster for **B)** *IRF1*, **C)** *MYC*, **D)** *E2F4*, and **E)** *NE2L2*. **F)** Violin
1022 plots of gene expression in superficial cluster 5 vs. all other superficial subclusters for SSc-specific
1023 differentially expressed genes that were uniquely expressed in cluster 5 within superficial EECs.
1024
1025 **Supplemental Figure 6. *FLI1* expression in human esophageal mucosa. A)** UMAP of integrated
1026 mucosal dataset, colored by *FLI1* expression. The distribution of *FLI1* expression by cell type is shown in
1027 the upper right. **B)** Distribution of *FLI1* expression by cell type and disease state. Endothelial cells had the
1028 highest *FLI1* expression, but the expression differences between diseases was not significant. **C)** Violin
1029 plot of the relative expression of FLI1 target genes in superficial cells.
1030
1031 **Supplemental Figure 7. Epithelial cell compartment proportions and additional correlations by**
1032 **motility phenotype. A-B)** Proportions of esophageal epithelial cell compartments by HRM **(A)** and FLIP
1033 **(B)** esophageal dysmotility phenotypes are shown for the proximal and distal esophagus samples. The
1034 bars denote the mean values, the vertical lines the standard deviations, and the points the individual
1035 sample proportions. **C)** The relationship between the proportion of superficial cells and the SSc disease
1036 duration is plotted for proximal and distal samples. Trendlines with 95% confidence intervals are plotted in
1037 pink, and correlation P-values are displayed in the upper right corners. **D)** The distribution of normalized
1038 gene expression values by esophageal motility phenotype are shown for the *TRIM11* gene for the FLIP,
1039 HRM, and NM esophageal dysmotility classifications. Unadjusted ordinal regression p-values are
1040 displayed in the upper left corners. *TRIM11* was the most strongly association gene with the dysmotility
1041 phenotypes by p-value. HRM, high-resolution manometry; FLIP, functional luminal imaging probe
1042 panometry; IEM, ineffective esophageal motility; EGJOO, esophagogastric junction outflow obstruction;
1043 NM, neuromyogenic model.

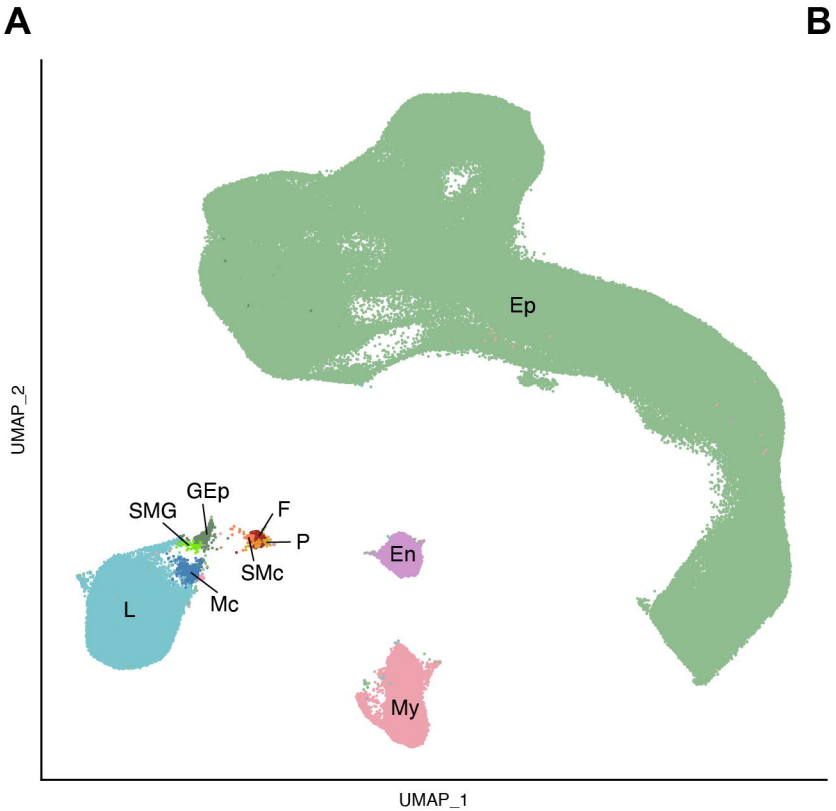
1044 **TABLES**

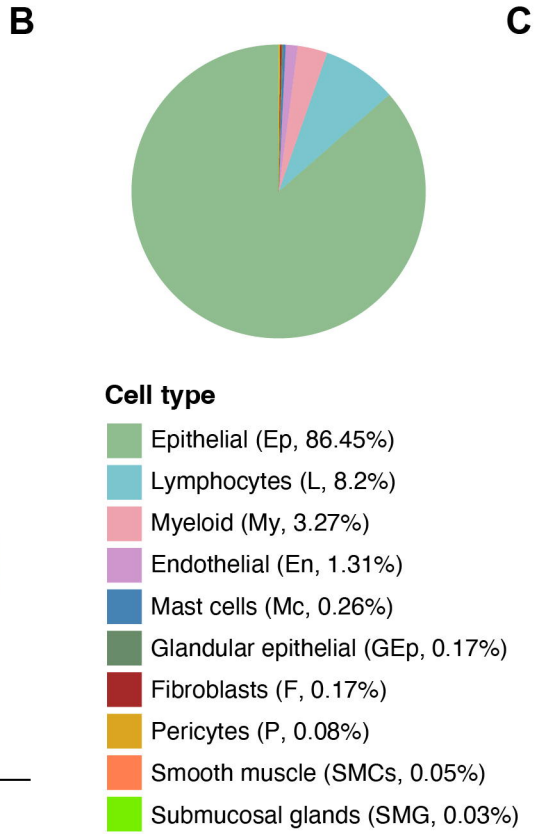
Table 1. Sc-RNA-seq patient demographics and clinical characteristics

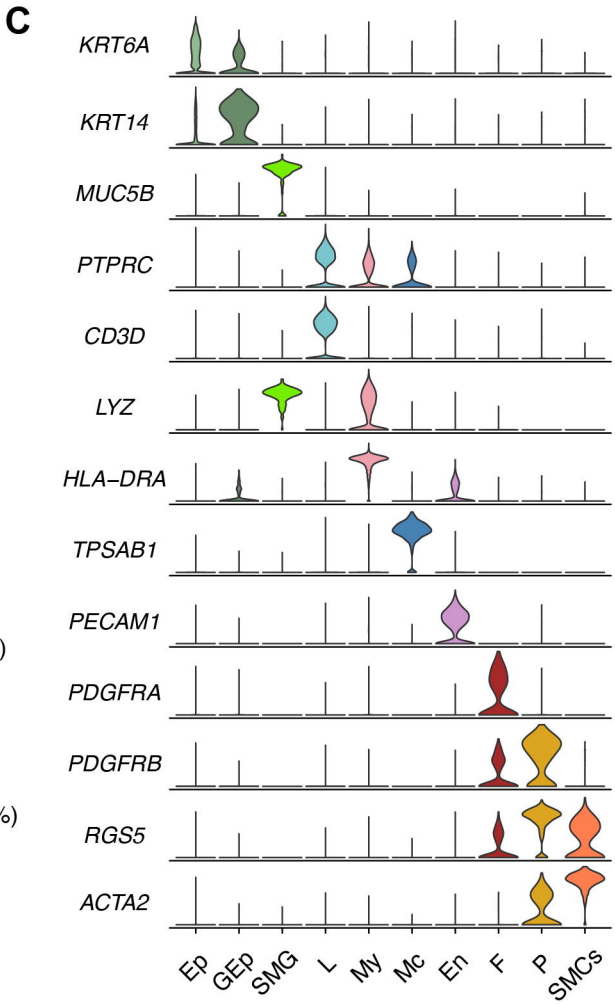
	HCs		SSc		GERD	
Demographics	**N = 6**		**N = 10**		**N = 4**	
Sex – female	5 (83.3%)		9 (90%)		3 (75%)	
Race – white	5 (83.3%)		8 (80%)		2 (50%)	
Ethnicity – non-Hispanic	5 (83.3%)		8 (80%)		2 (50%)	
Race/ethnicity – unknown/NR	1 (16.7%)		2 (20%)		2 (50%)	
Traits	**N**	**Median (IQR)**	**N**	**Median (IQR)**	**N**	**Median (IQR)**
Age (yrs)	6	27 (25-29)	10	54 (47-70)	4	42 (30-56)
BMI (kg/m^2)	6	23 (22-25)	10	22 (19-30)	4	24 (23-36)
Esophageal motility						
Mean DCI (mmHg•s•cm)	6	464.2 (386-6-1026.1)	10	57.5 (0.0-206.1)	2	509.6 (367.4-651.7)
Pressure 60ml (mmHg)	6	43.5 (38.3-54.5)	9	31.0 (30.5-40.3)	0	-
EGJ barrier function						
End-expiratory EGJ pressure (mmHg)	6	7 (4-8)	10	6 (3-10)	2	12 (11-12)
EGJ contractile index (mmHg•cm)	6	19.4 (15.2-29.0)	10	21.0 (19.1-27.5)	2	19.0 (10.8-27.1)
Median IRP (mmHg)	6	4 (3-9)	10	6 (4-12)	2	8 (6-9)
EGJ-DI (mm^2/mmHg)	6	6.2 (5.2-8.5)	9	6.3 (4.1-9.2)	-	-
Reflux severity						
AET (%)	6	0.8 (0.5-2.8)	4	8.7 (3.3-14.5)	3	16.8 (13.9-17.3)
GerdQ[118] score	6	0 (0-0)	10	3 (1-4)	3	3 (3-4)
BEDQ[38] score	6	0 (0-0)	8	8 (3-14)	3	1 (1-1)
NEQOL[39] score	6	56 (56-56)	8	31 (23-47)	3	36 (29-46)

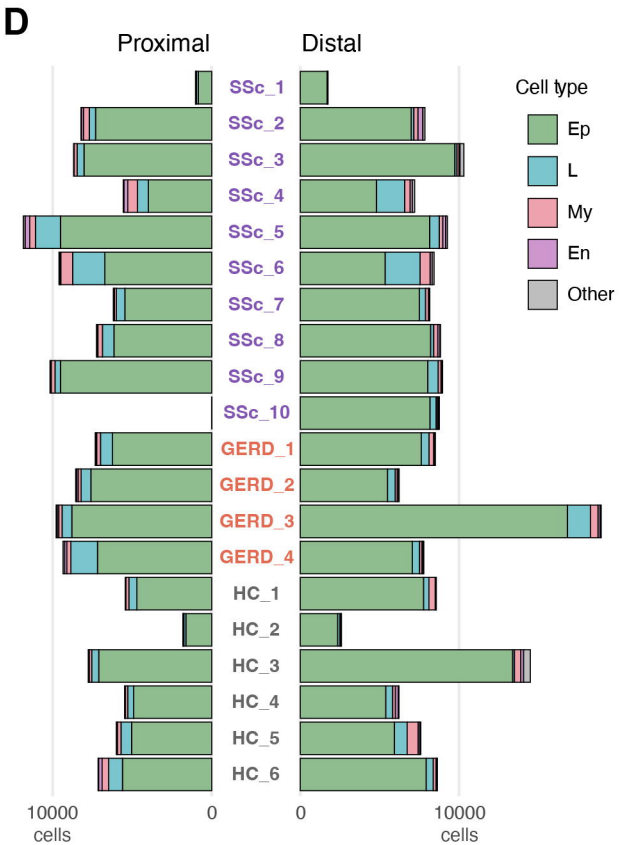
1045 AET, acid exposure time; BEDQ, brief esophageal dysphagia questionnaire; BMI, body mass index; DCI, distal contractile integral;
1046 DI, distensibility index; EGJ, esophagogastric junction; HCs, healthy controls; IQR, inter-quartile range; IRP, integrated relaxation
1047 pressure; NEQOL, Northwestern esophageal quality of life; NR, not reported; scRNA-seq, single-cell RNA sequencing.

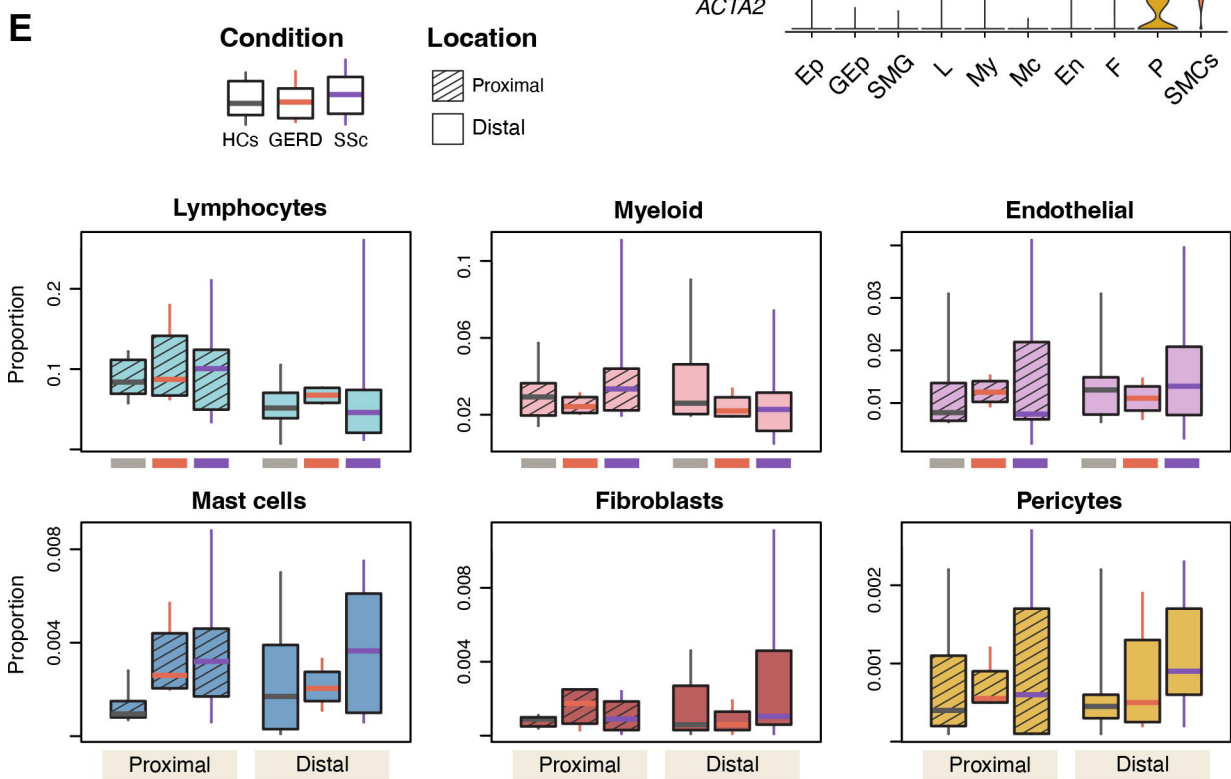
1048

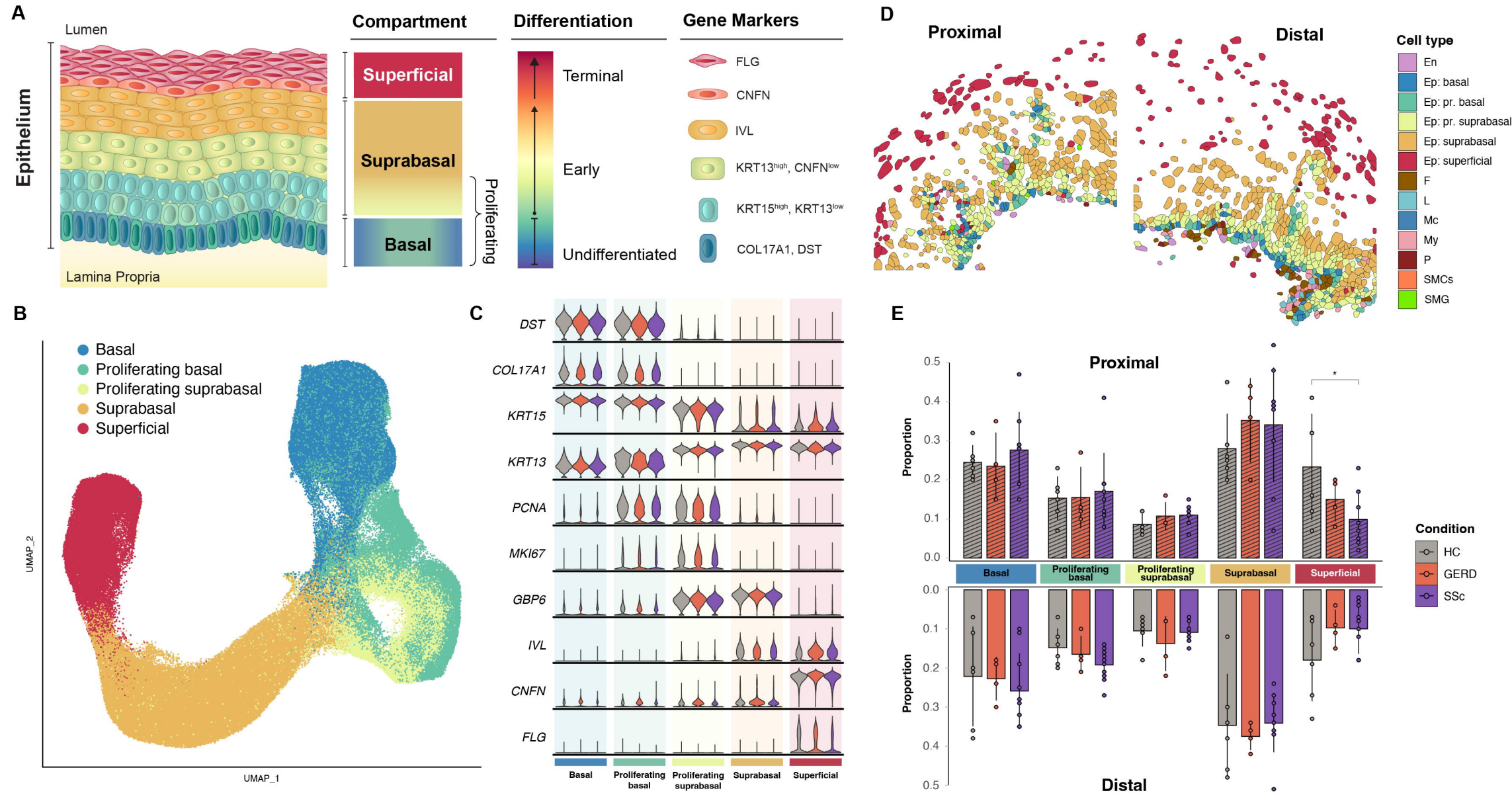

A
Ep
GEp
SMG
F
P
SMc
Mc
L
En
My
UMAP_2
UMAP_1

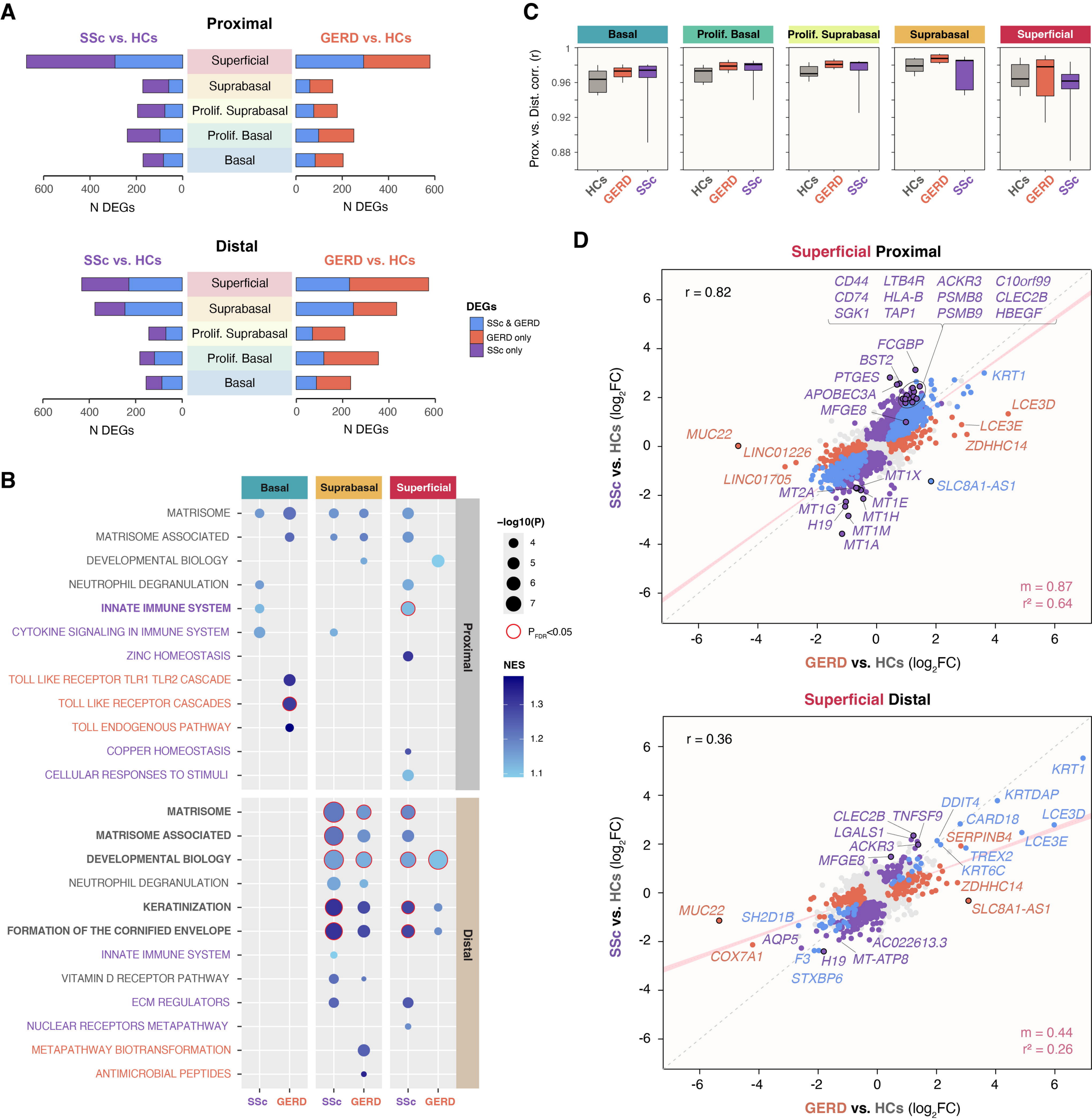

B
Cell type
Epithelial (Ep, 86.45%)
Lymphocytes (L, 8.2%)
Myeloid (My, 3.27%)
Endothelial (En, 1.31%)
Mast cells (Mc, 0.26%)
Glandular epithelial (GEp, 0.17%)
Fibroblasts (F, 0.17%)
Pericytes (P, 0.08%)
Smooth muscle (SMCs, 0.05%)
Submucosal glands (SMG, 0.03%)

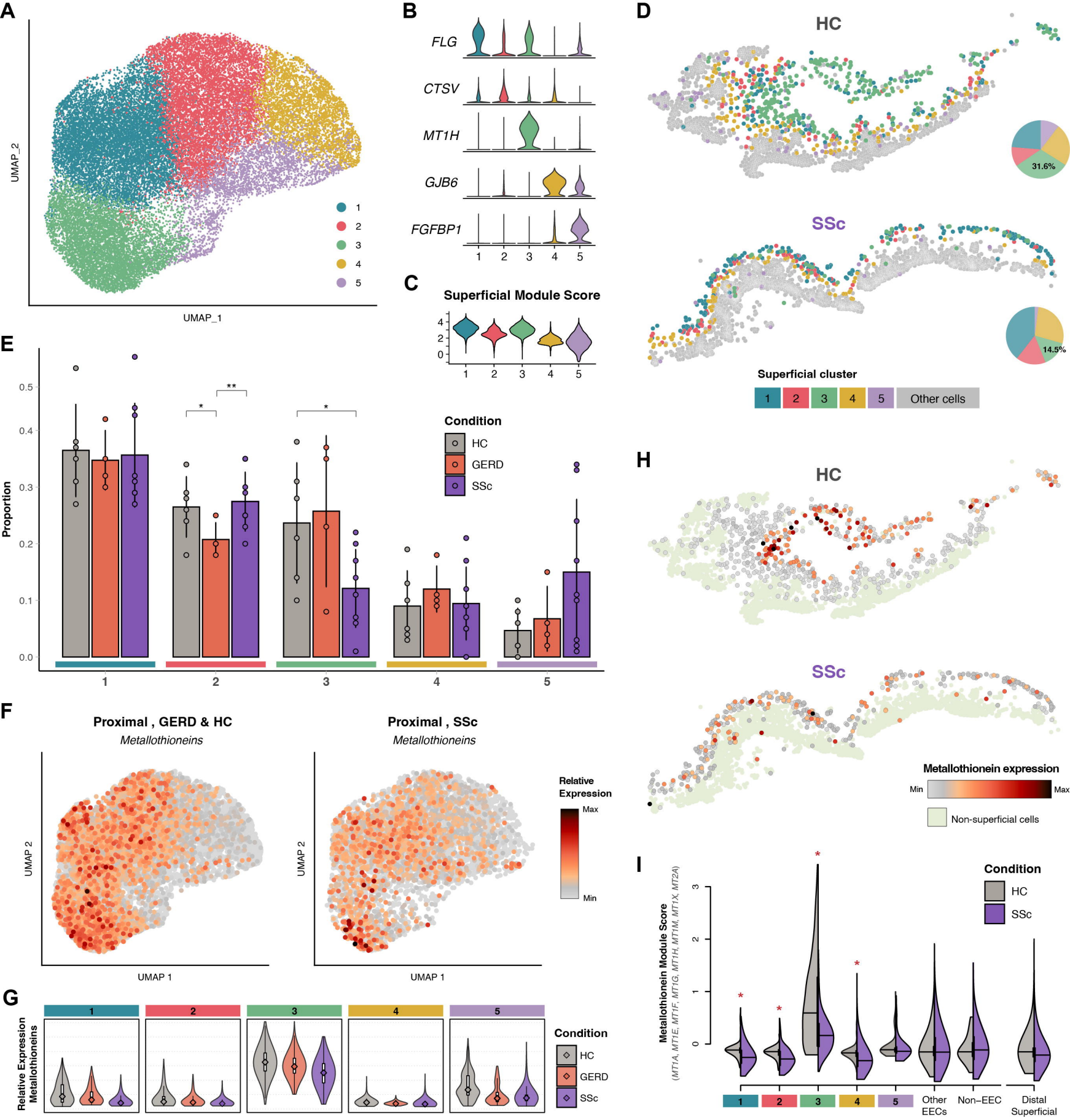

C
KRT6A
KRT14
MUC5B
PTPRC
CD3D
LYZ
HLA–DRA
TPSAB1
PECAM1
PDGFRA
PDGFRB
RGS5
ACTA2
Ep
GEp
SMG
L
My
Mc
En
F
P
SMCs

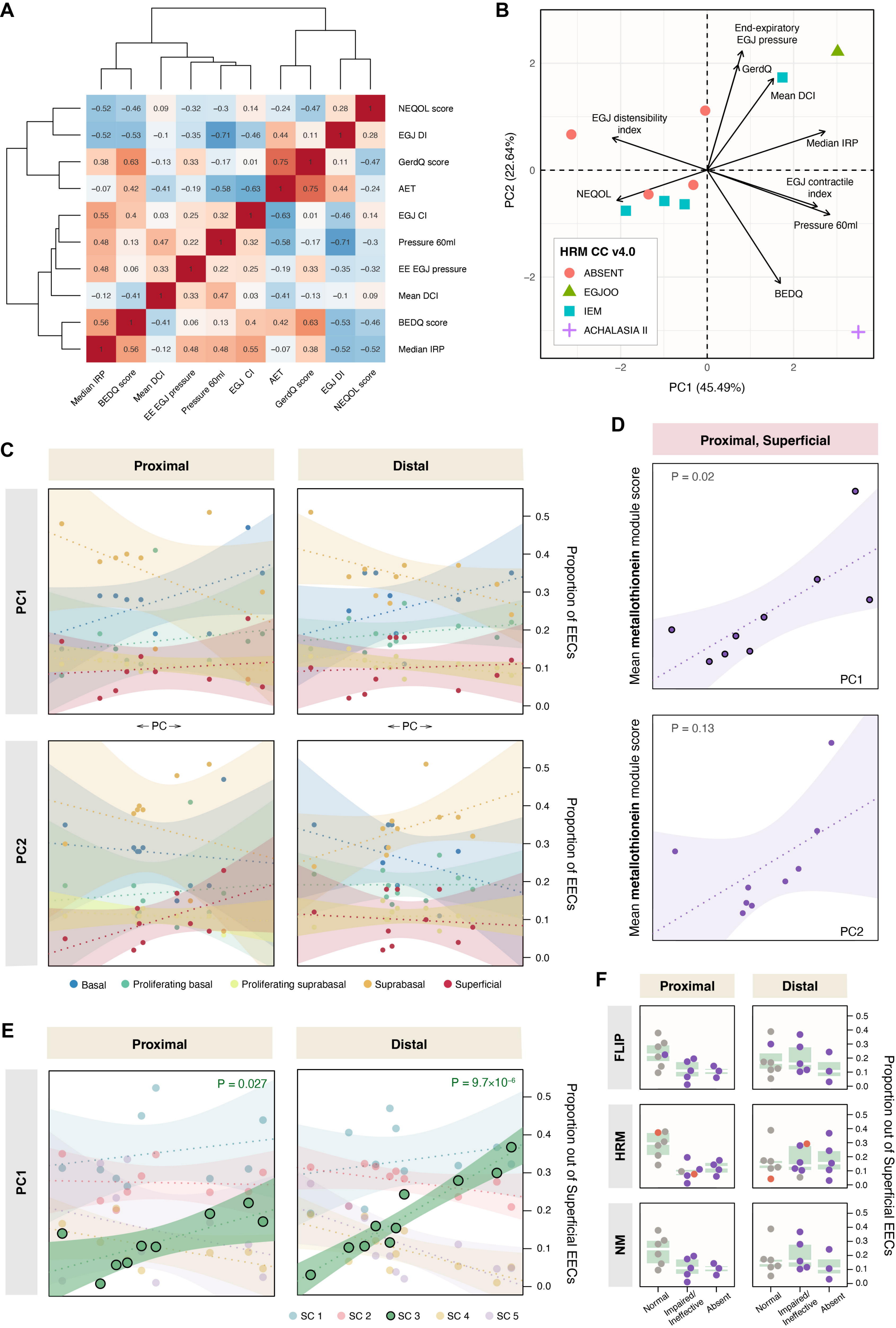

D
Proximal
Distal
SSc_1
SSc_2
SSc_3
SSc_4
SSc_5
SSc_6
SSc_7
SSc_8
SSc_9
SSc_10
GERD_1
GERD_2
GERD_3
GERD_4
HC_1
HC_2
HC_3
HC_4
HC_5
HC_6
Cell type
Ep
L
My
En
Other
10000 cells
0
0
10000 cells

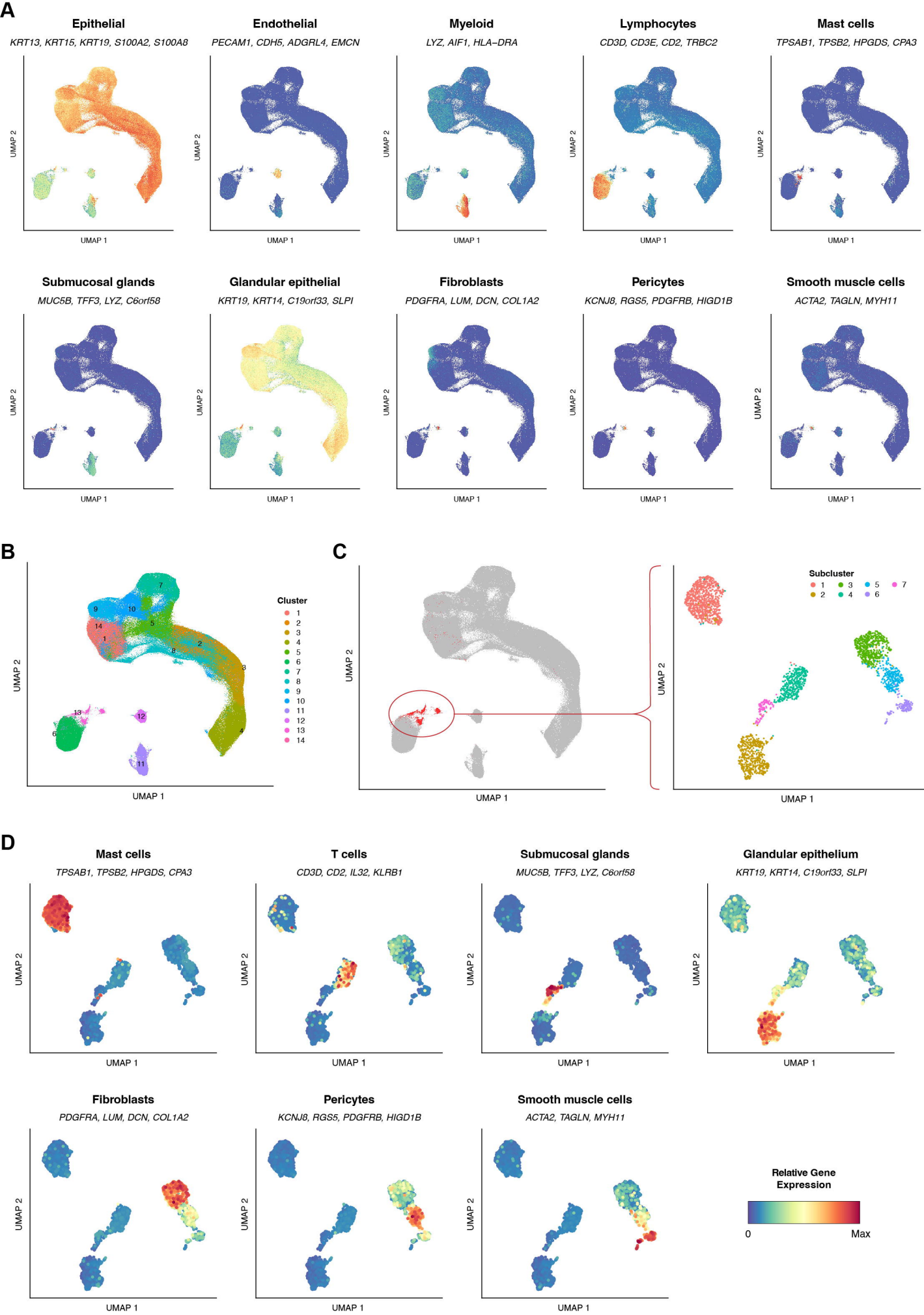

E
Condition
HCs
GERD
SSc
Location
Proximal
Distal
Lymphocytes
Myeloid
Endothelial
Mast cells
Fibroblasts
Pericytes
Proportion
Proximal
Distal

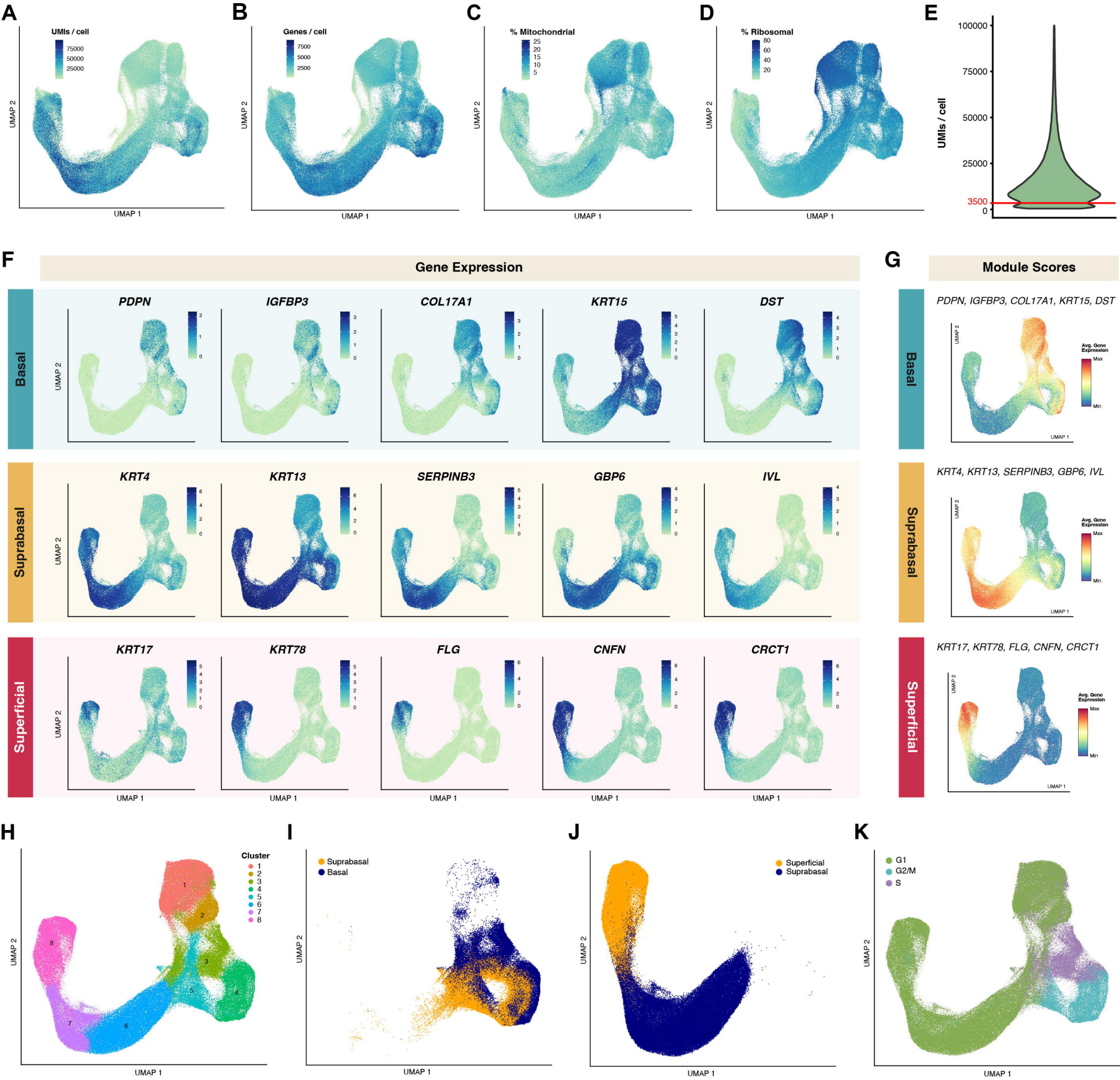

A
Lumen
Epithelium
Lamina Propria
Compartment
Superficial
Suprabasal
Basal
Proliferating
Differentiation
Terminal
Early
Undifferentiated
Gene Markers
FLG
CNFN
IVL
KRT13high, CNFNlow
KRT15high, KRT13low
COL17A1, DST
B
Basal
Proliferating basal
Proliferating suprabasal
Suprabasal
Superficial
UMAP_2
UMAP_1
C
DST
COL17A1
KRT15
KRT13
PCNA
MKI67
GBP6
IVL
CNFN
FLG
Basal
Proliferating basal
Proliferating suprabasal
Suprabasal
Superficial
D
Proximal
Distal
Cell type
En
Ep: basal
Ep: pr. basal
Ep: pr. suprabasal
Ep: suprabasal
Ep: superficial
F
L
Mc
My
P
SMCs
SMG
E
Proximal
Distal
Proportion
0.0
0.1
0.2
0.3
0.4
0.5
*
Basal
Proliferating basal
Proliferating suprabasal
Suprabasal
Superficial
Condition
HC
GERD
SSc

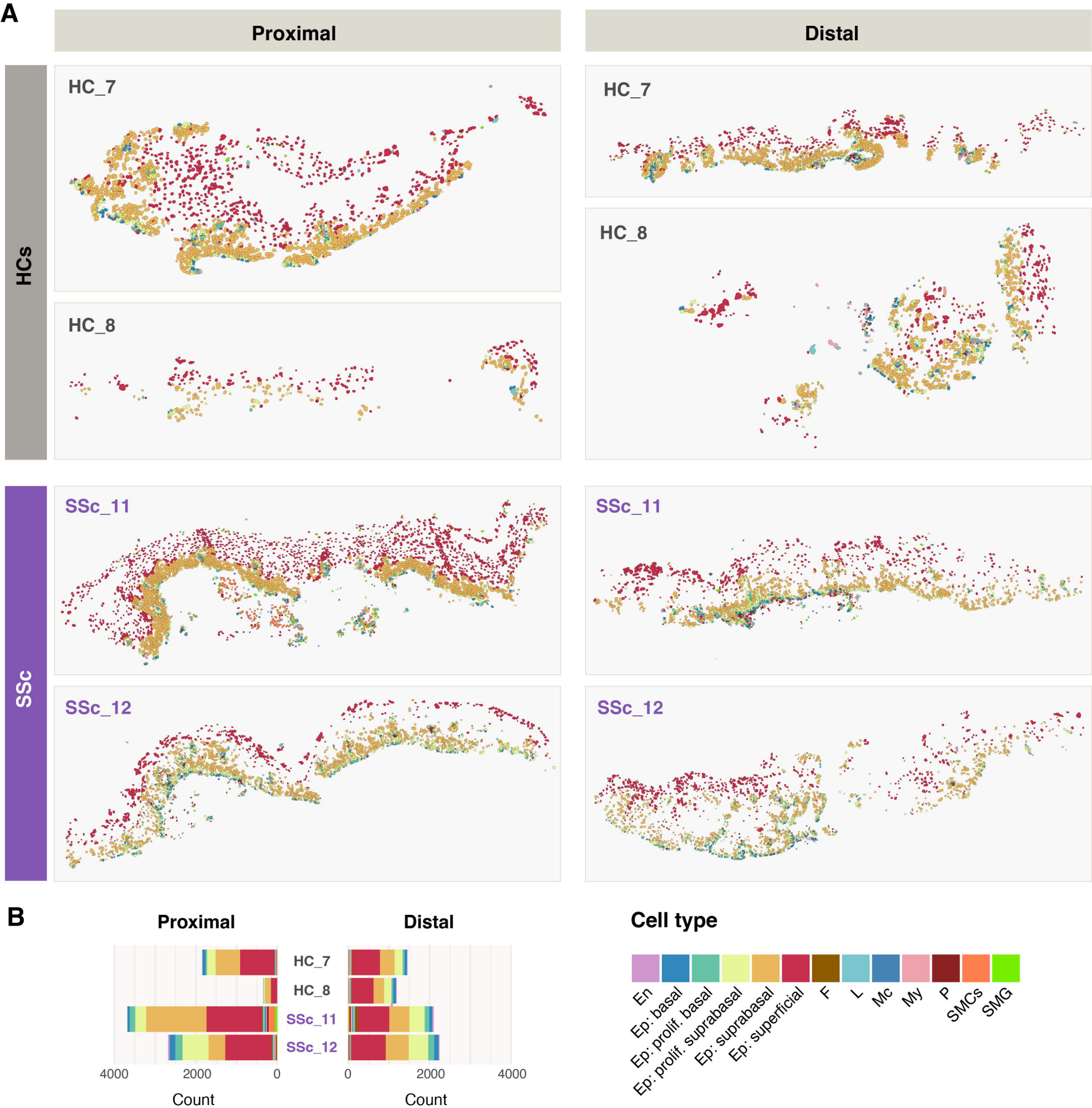

A
Proximal
SSc vs. HCs
GERD vs. HCs
Superficial
Suprabasal
Prolif. Suprabasal
Prolif. Basal
Basal
600 400 200 0
0 200 400 600
N DEGs
Distal
DEGs
SSc & GERD
GERD only
SSc only
B
Basal
Suprabasal
Superficial
MATRISOME
MATRISOME ASSOCIATED
DEVELOPMENTAL BIOLOGY
NEUTROPHIL DEGRANULATION
INNATE IMMUNE SYSTEM
CYTOKINE SIGNALING IN IMMUNE SYSTEM
ZINC HOMEOSTASIS
TOLL LIKE RECEPTOR TLR1 TLR2 CASCADE
TOLL LIKE RECEPTOR CASCADES
TOLL ENDOGENOUS PATHWAY
COPPER HOMEOSTASIS
CELLULAR RESPONSES TO STIMULI
Proximal
KERATINIZATION
FORMATION OF THE CORNIFIED ENVELOPE
VITAMIN D RECEPTOR PATHWAY
ECM REGULATORS
NUCLEAR RECEPTORS METAPATHWAY
METAPATHWAY BIOTRANSFORMATION
ANTIMICROBIAL PEPTIDES
Distal
SSc GERD SSc GERD SSc GERD
−log10(P)
4
5
6
7
P_FDR<0.05
NES
1.3
1.2
1.1
C
Prox. vs. Dist. corr. (r)
1
0.96
0.92
0.88
HCs GERD SSc
D
Superficial Proximal
r = 0.82
CD44 LTB4R ACKR3 C10orf99
CD74 HLA-B PSMB8 CLEC2B
SGK1 TAP1 PSMB9 HBEGF
FCGBP
BST2
PTGES
APOBEC3A
MFGE8
KRT1
LCE3D
LCE3E
ZDHHC14
MUC22
LINC01226
LINC01705
MT2A
MT1X
SLC8A1-AS1
MT1G
MT1E
MT1H
H19
MT1M
MT1A
m = 0.87
r² = 0.64
SSc vs. HCs (log2FC)
GERD vs. HCs (log2FC)
Superficial Distal
r = 0.36
KRT1
KRTDAP
DDIT4
CLEC2B TNFSF9
CARD18
LCE3D
LGALS1
SERPINB4
LCE3E
ACKR3
TREX2
MFGE8
KRT6C
ZDHHC14
MUC22
SH2D1B
SLC8A1-AS1
AQP5
AC022613.3
COX7A1
F3
H19
MT-ATP8
STXBP6
m = 0.44
r² = 0.26
SSc vs. HCs (log2FC)
GERD vs. HCs (log2FC)

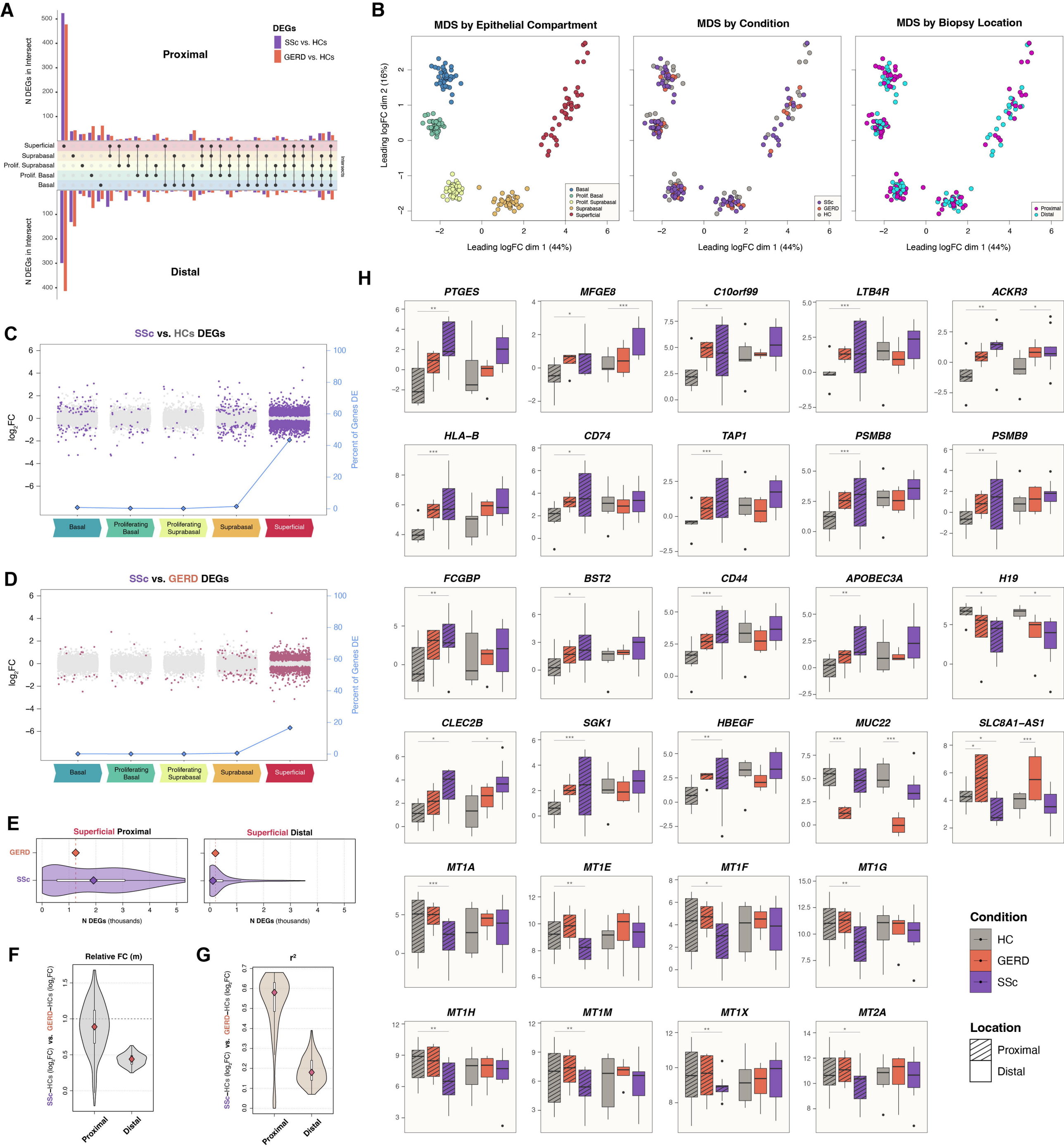

A
UMAP_2
UMAP_1
1
2
3
4
5
B
FLG
CTSV
MT1H
GJB6
FGFBP1
1 2 3 4 5
C
Superficial Module Score
4
2
0
1 2 3 4 5
D
HC
31.6%
SSc
14.5%
Superficial cluster
1 2 3 4 5 Other cells
E
Proportion
0.5
0.4
0.3
0.2
0.1
0.0
*
**
*
Condition
HC
GERD
SSc
1 2 3 4 5
F
Proximal , GERD & HC
Metallothioneins
Proximal , SSc
Metallothioneins
UMAP 2
UMAP 1
Relative Expression
Max
Min
G
1 2 3 4 5
Relative Expression Metallothioneins
Condition
HC
GERD
SSc
H
HC
SSc
Metallothionein expression
Min
Max
Non-superficial cells
I
Metallothionein Module Score
(MT1A, MT1E, MT1F, MT1G, MT1H, MT1M, MT1X, MT2A)
Condition
HC
SSc
3
2
1
0
1 2 3 4 5 Other EECs Non–EEC Distal Superficial

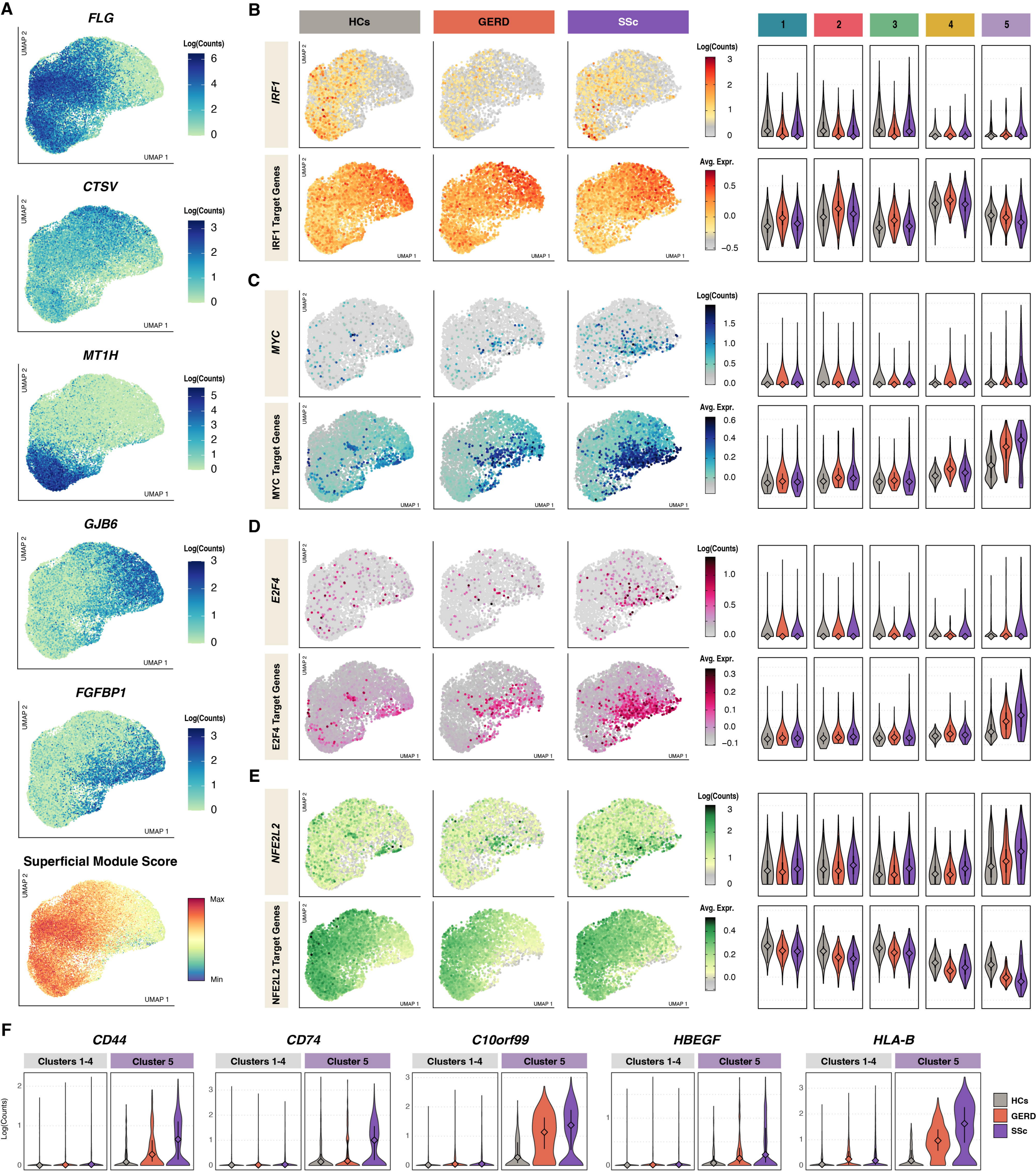

A
NEQOL score
EGJ DI
GerdQ score
AET
EGJ CI
Pressure 60ml
EE EGJ pressure
Mean DCI
BEDQ score
Median IRP
B
End-expiratory EGJ pressure
GerdQ
Mean DCI
EGJ distensibility index
Median IRP
EGJ contractile index
NEQOL
Pressure 60ml
BEDQ
HRM CC v4.0
ABSENT
EGJOO
IEM
ACHALASIA II
PC2 (22.64%)
PC1 (45.49%)
C
Proximal
Distal
PC1
PC2
← PC →
Proportion of EECs
Basal
Proliferating basal
Proliferating suprabasal
Suprabasal
Superficial
D
Proximal, Superficial
P = 0.02
P = 0.13
Mean metallothionein module score
PC1
PC2
E
P = 0.027
P = 9.7×10⁻⁶
Proportion out of Superficial EECs
SC 1
SC 2
SC 3
SC 4
SC 5
F
FLIP
HRM
NM
Normal
Impaired/ Ineffective
Absent

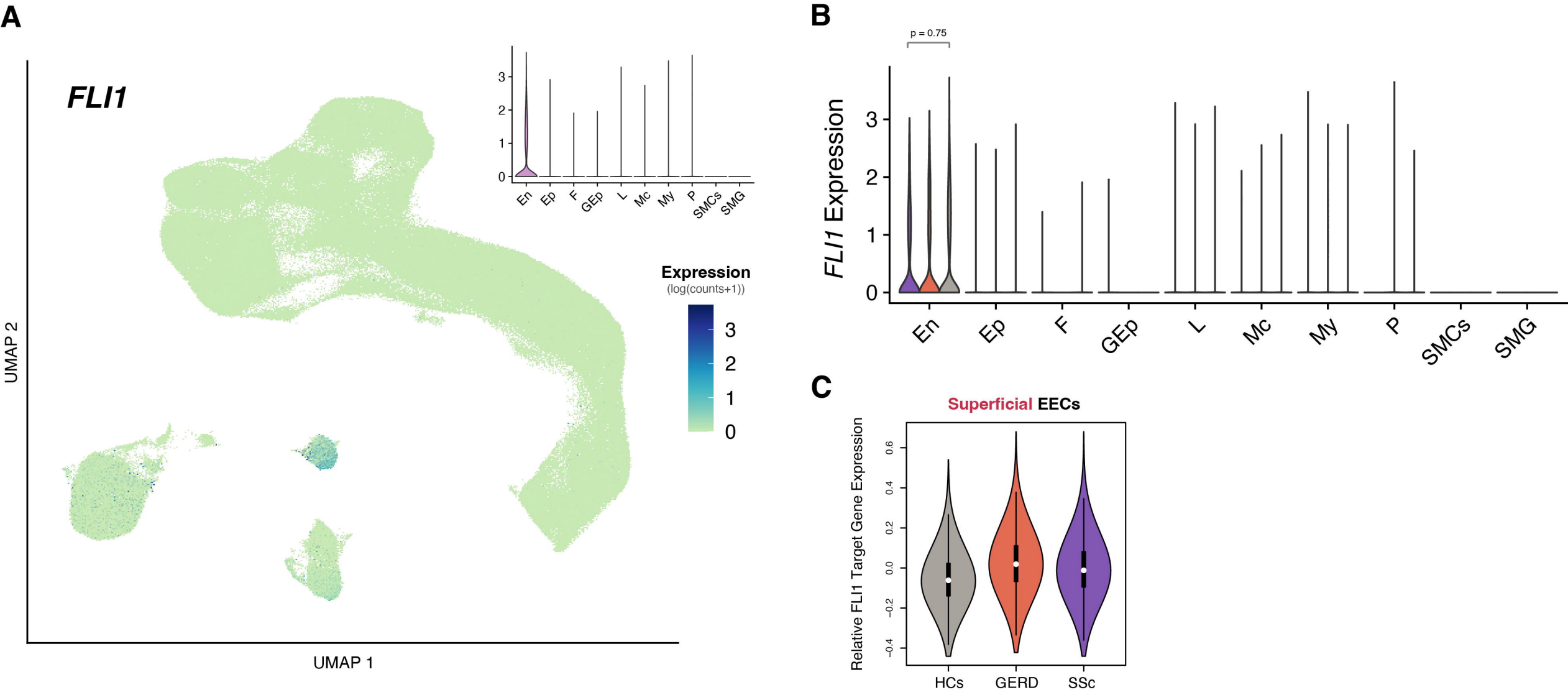

A
Epithelial
KRT13, KRT15, KRT19, S100A2, S100A8
Endothelial
PECAM1, CDH5, ADGRL4, EMCN
Myeloid
LYZ, AIF1, HLA–DRA
Lymphocytes
CD3D, CD3E, CD2, TRBC2
Mast cells
TPSAB1, TPSB2, HPGDS, CPA3
Submucosal glands
MUC5B, TFF3, LYZ, C6orf58
Glandular epithelial
KRT19, KRT14, C19orf33, SLPI
Fibroblasts
PDGFRA, LUM, DCN, COL1A2
Pericytes
KCNJ8, RGS5, PDGFRB, HIGD1B
Smooth muscle cells
ACTA2, TAGLN, MYH11
UMAP 1
UMAP 2
B
Cluster
1
2
3
4
5
6
7
8
9
10
11
12
13
14
C
Subcluster
1
2
3
4
5
6
7
D
Mast cells
TPSAB1, TPSB2, HPGDS, CPA3
T cells
CD3D, CD2, IL32, KLRB1
Submucosal glands
MUC5B, TFF3, LYZ, C6orf58
Glandular epithelium
KRT19, KRT14, C19orf33, SLPI
Fibroblasts
PDGFRA, LUM, DCN, COL1A2
Pericytes
KCNJ8, RGS5, PDGFRB, HIGD1B
Smooth muscle cells
ACTA2, TAGLN, MYH11
Relative Gene Expression
0
Max

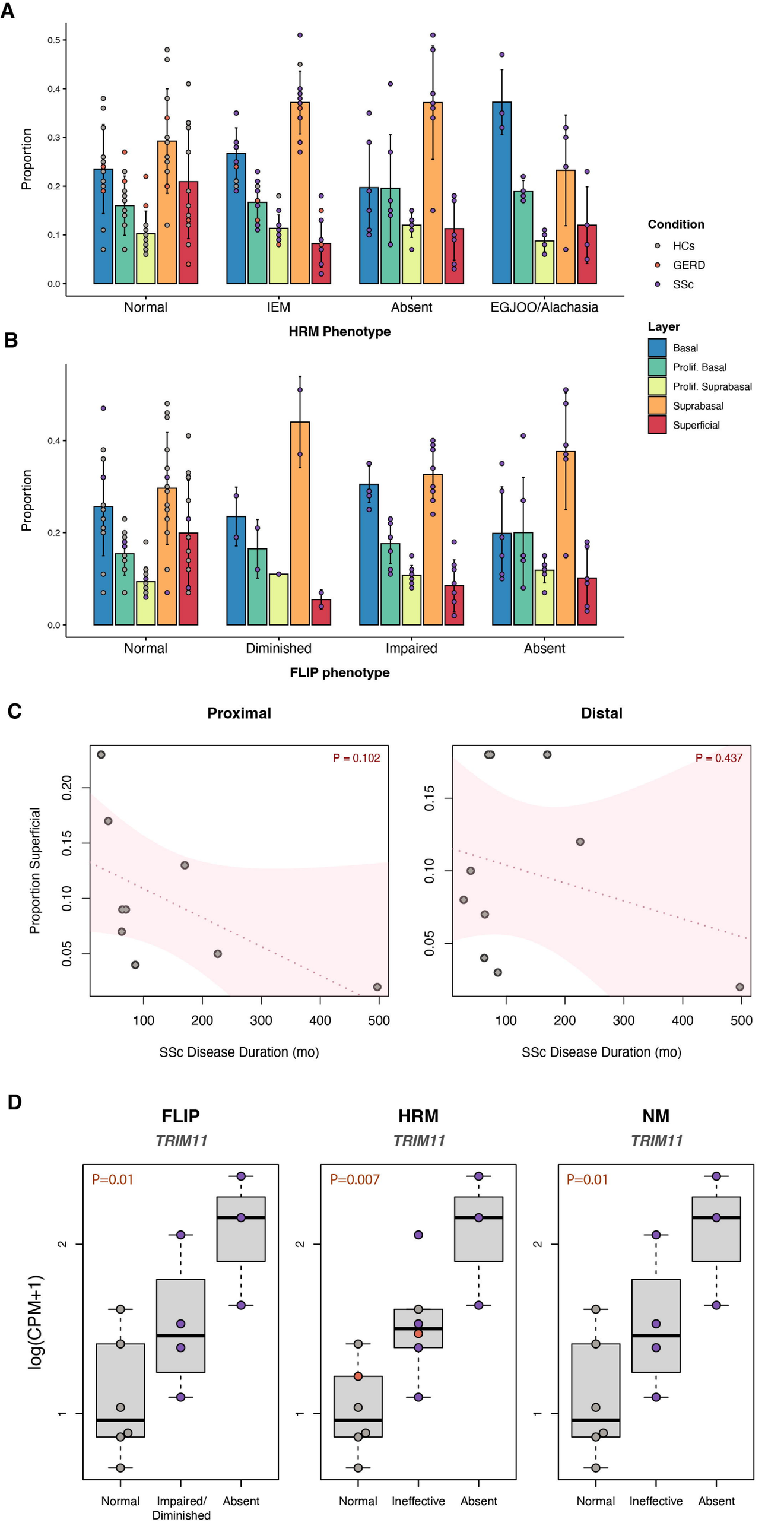

A
UMIs / cell
75000
50000
25000
UMAP 2
UMAP 1
B
Genes / cell
7500
5000
2500
C
% Mitochondrial
25
20
15
10
5
D
% Ribosomal
80
60
40
20
E
UMIs / cell
100000
75000
50000
25000
3500
0
F
Gene Expression
Basal
PDPN
IGFBP3
COL17A1
KRT15
DST
Suprabasal
KRT4
KRT13
SERPINB3
GBP6
IVL
Superficial
KRT17
KRT78
FLG
CNFN
CRCT1
G
Module Scores
PDPN, IGFBP3, COL17A1, KRT15, DST
KRT4, KRT13, SERPINB3, GBP6, IVL
KRT17, KRT78, FLG, CNFN, CRCT1
Avg. Gene Expression
Max
Min
H
Cluster
1
2
3
4
5
6
7
8
I
Suprabasal
Basal
J
Superficial
Suprabasal
K
G1
G2/M
S


A
Proximal
Distal
HCs
HC_7
HC_8
SSc
SSc_11
SSc_12
B
Proximal
Distal
HC_7
HC_8
SSc_11
SSc_12
4000
2000
0
Count
Cell type
En
Ep: basal
Ep: prolif. basal
Ep: prolif. suprabasal
Ep: suprabasal
Ep: superficial
F
L
Mc
My
P
SMCs
SMG


A
Proximal
Distal
DEGs
SSc vs. HCs
GERD vs. HCs
N DEGs in Intersect
Superficial
Suprabasal
Prolif. Suprabasal
Prolif. Basal
Basal
Intersects
B
MDS by Epithelial Compartment
MDS by Condition
MDS by Biopsy Location
Leading logFC dim 1 (44%)
Leading logFC dim 2 (16%)
Basal
Prolif. Basal
Prolif. Suprabasal
Suprabasal
Superficial
SSc
GERD
HC
Proximal
Distal
C
SSc vs. HCs DEGs
log2FC
Percent of Genes DE
Basal
Proliferating Basal
Proliferating Suprabasal
Suprabasal
Superficial
D
SSc vs. GERD DEGs
E
Superficial Proximal
Superficial Distal
GERD
SSc
N DEGs (thousands)
F
Relative FC (m)
SSc–HCs (log2FC) vs. GERD–HCs (log2FC)
Proximal
Distal
G
r2
H
PTGES
MFGE8
C10orf99
LTB4R
ACKR3
HLA–B
CD74
TAP1
PSMB8
PSMB9
FCGBP
BST2
CD44
APOBEC3A
H19
CLEC2B
SGK1
HBEGF
MUC22
SLC8A1–AS1
MT1A
MT1E
MT1F
MT1G
MT1H
MT1M
MT1X
MT2A
Condition
HC
GERD
SSc
Location
Proximal
Distal


A
FLG
CTSV
MT1H
GJB6
FGFBP1
Superficial Module Score
UMAP 1
UMAP 2
Log(Counts)
Max
Min
B
HCs
GERD
SSc
1
2
3
4
5
IRF1
IRF1 Target Genes
Avg. Expr.
C
MYC
MYC Target Genes
D
E2F4
E2F4 Target Genes
E
NFE2L2
NFE2L2 Target Genes
F
CD44
CD74
C10orf99
HBEGF
HLA-B
Clusters 1-4
Cluster 5
Log(Counts)
HCs
GERD
SSc


A
FLI1
UMAP 2
UMAP 1
Expression
(log(counts+1))
3
2
1
0
En
Ep
F
GEp
L
Mc
My
P
SMCs
SMG
B
p = 0.75
FLI1 Expression
C
Superficial EECs
Relative FLI1 Target Gene Expression
HCs
GERD
SSc


A
Proportion
0.0
0.1
0.2
0.3
0.4
0.5
Normal
IEM
Absent
EGJOO/Alachasia
HRM Phenotype
Condition
HCs
GERD
SSc
B
Proportion
0.0
0.2
0.4
Normal
Diminished
Impaired
Absent
FLIP phenotype
Layer
Basal
Prolif. Basal
Prolif. Suprabasal
Suprabasal
Superficial
C
Proximal
Distal
P = 0.102
P = 0.437
Proportion Superficial
0.05
0.10
0.15
0.20
100
200
300
400
500
SSc Disease Duration (mo)
D
FLIP
HRM
NM
TRIM11
P=0.01
P=0.007
P=0.01
log(CPM+1)
1
2
Normal
Impaired/ Diminished
Absent
Ineffective