The esophageal epithelium in systemic sclerosis: cellular and molecular dysregulation revealed by single-cell RNA sequencing

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

Affiliations:

1. Division of Rheumatology, Department of Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL, USA
2. Division of Gastroenterology & Hepatology, Department of Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL, USA
3. Kenneth C. Griffin Esophageal Center of Northwestern Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL, USA
4. Department of Preventive Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL, USA
5. 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 progressive 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 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 outermost, 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), 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. 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.

45 **INTRODUCTION**

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

55 There are multiple, interrelated mechanisms that may cause esophageal dysmotility in SSc.
56 Vascular damage and neurogenic impairment lead to smooth muscle cell fibrosis[13], but esophageal muscle
57 atrophy has also been observed in SSc without vasculopathy or fibrosis[14]. Inflammatory signatures[15],
58 absence of anti-centromere antibodies[16], and presence of anti-topoisomerase I antibodies are associated with
59 esophageal complications, as well[8]. Complicating research efforts is significant clinical and molecular
60 heterogeneity observed between SSc patients[17-19] and difficulty distinguishing between effects from
61 autoimmune processes and secondary factors such as severe reflux. Consequently, the pathogenesis of
62 esophageal disease in SSc remains poorly understood.

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

70 esophagus persist in mice additionally lacking mature T and B cells, suggesting that the fibrosis is
71 principally driven by the epithelia[20,24].

72 Most studies implicating epithelial cells in SSc esophageal dysmotility have been performed in
73 mice or *in vitro*, and molecular analyses in humans have so far been limited to bulk tissues[15]. Furthermore,
74 single-cell gene expression studies performed in other diseases of the esophagus, including squamous cell
75 carcinoma[25,26] and allergic eosinophilic esophagitis[27,28], have identified substantial cellular and molecular
76 changes in EECs compared to healthy tissue, such as significant expansion of non-proliferative suprabasal
77 cells[27,28]. Therefore, to comprehensively investigate the effects of SSc on the esophageal epithelium, we
78 performed single-cell RNA sequencing (scRNA-seq) of esophageal mucosal biopsies from SSc patients
79 with clinically significant esophageal involvement and compared EEC distributions and gene expression
80 signatures to individuals with gastroesophageal reflux disease (GERD) and healthy controls (HCs). By
81 examining gene expression profiles from human tissue at single-cell resolution, this work sheds essential
82 light on the cellular roots of esophageal dysfunction in SSc by clarifying the pathogenic role of the
83 squamous epithelium, one of the most integral tissues supporting healthy esophageal function.

84

85 **METHODS**

86

87 **Subjects**

88 The SSc study population were recruited from adult patients aged 18 to 89 diagnosed with SSc according
89 to the 2013 ACR/EULAR classification criteria, who visited the Northwestern Medicine Esophageal Center
90 for esophageal symptom evaluation and motility testing, as described previously[29], between 2020 and 2021.
91 Patients with GERD were recruited at the primary visit following positive wireless pH testing[28]. Patients
92 with technically inadequate panometry or manometry studies, previous foregut surgeries (including prior
93 pneumatic dilation), or mechanical obstructions in the esophagus such as esophageal stricture, eosinophilic
94 esophagitis, severe reflux esophagitis (Los Angeles classification C or D), or hiatal hernias larger than 5
95 cm were excluded. HCs met asymptomatic criteria including lack of esophageal symptoms, no history of

96 alcohol dependency or tobacco use, body mass index less than 30kg/m^2, and no current antacid or proton
97 pump inhibitor treatment[28]. The study protocol received approval from the Northwestern University
98 Institutional Review Board.
99
100 **Symptom evaluation and motility testing**
101 Esophageal function testing included functional lumen imaging probe (FLIP) panometry during sedated
102 endoscopy, high-resolution manometry (HRM), and in some subjects, 24-hour pH-impedance or timed-
103 barium esophagram, which were completed and interpreted as previously described[30-33]. HRM
104 classifications were based on the application of the Chicago Classification v4.0 to 10 supine and 5 upright
105 test swallows[34]. FLIP panometry classifications were based on previous evaluation of asymptomatic
106 volunteers and patients[35-37]. During endoscopy (same encounter as FLIP), esophageal mucosal biopsies
107 were obtained at approximately 5-cm ("distal esophagus") and 15-cm ("proximal esophagus") proximal to
108 the squamocolumnar junction. Acid exposure was measured as the percentage of time with esophageal
109 pH<4 over a monitoring period of 24 hours. Most patients also completed validated symptom-
110 questionnaires (patient-reported outcomes) on the day of esophageal motility testing with FLIP or HRM,
111 including the Brief Esophageal Dysphagia Questionnaire (BEDQ), the Gastroesophageal Reflux Disease
112 Questionnaire (GerdQ), and the Northwestern esophageal quality of life (NEQOL) survey.[38-40].
113
114 **Sample processing and sequencing**
115 Esophageal mucosal biopsies from proximal and distal esophagus were processed for scRNA-seq as
116 described in Clevenger et al[28]. Briefly, following tissue digestion and filtering, cell suspensions met 85%
117 minimum viability via Cellometer Auto2000 (Nexcelom Bioscience). Cells were loaded into a Chromium
118 iX controller (10X Genomics) on a Chromium Next GEM Chip G (10X Genomics) to capture
119 approximately 10,000 cells per sample and were processed for encapsulation following the manufacturer's
120 protocol. Cell barcoding and library construction were performed using the 10X Genomics Chromium
121 Next GEM Single Cell 3' Reagents Kits v3.1 and Dual Index Kit TT Set A according to the manufacturer's

122 protocol. Sequencing was performed by the NUseq Core using the Illumina Novaseq 6000. Paired-end
123 reads consisting of a 28-base-pair read for cell barcodes and unique molecular identifiers and a 90-base-
124 pair read for transcripts were aligned to the GRCh38 reference transcriptome using Cell Ranger v6.1.2.
125

126 **Sample and cell quality control**

127 Samples with high ambient RNA (<70% of sequencing reads mapped to cells) were removed from
128 consideration. Gene expression counts generated by Cell Ranger were analyzed using Seurat v4.3.0[41] in R
129 v4.3.0. Cells with low gene diversity (<200 unique genes), low unique molecular identifier counts (<500),
130 or excessive counts (>100,000) were removed from consideration. Cells with high proportions of
131 mitochondrial DNA reads, indicative of apoptosis or lysis, were removed using the bivariate regression
132 models implemented in the miQC package[42]. Cells were filtered if their proportion of reads mapped to
133 mitochondrial DNA exceeded 0.05 and their posterior probability of belonging to a compromised cell
134 distribution was greater than 0.75 according to miQC. Doublets, in which more than one cell is clumped
135 together in a single droplet, were predicted and removed using scDblFinder[43] with an expected, additive
136 doublet rate of 0.01 per thousand cells with standard deviation of 0.01.
137

138 **Sample integration and cell type annotation**

139 Individual samples were normalized and integrated on 3,000 variable features using Seurat's SCTransform
140 procedure[44] with v2 regularization[45] and canonical correlation analysis dimension reduction. Integration
141 anchors were determined using a reference consisting of three pairs of proximal and distal samples from
142 HCs. Principal component (PC) analysis was performed on the transformed and integrated expression
143 counts. Uniform manifold approximation and projection (UMAP)[46] embeddings were calculated on the top
144 40 PCs with n=200 neighbors. Cell clustering was performed using Seurat's modularity-based clustering
145 on a shared nearest neighbor (SNN) graph with k=25 on the top 40 PCs. Clusters were annotated according
146 to the expression of established cell-specific gene markers (**Supplemental Table 1**, **Supplemental Figure**

147 **1**). For annotation of smaller cell clusters, we first filtered out the epithelial, myeloid, lymphoid, and
148 endothelial cell clusters and then recomputed PCs and re-clustered the subset (PCs = 25, k=15).
149
150 **Epithelial cell characterization**
151 After isolating the subset of epithelial cell clusters, we further filtered cells with disproportionately low
152 read counts (≤3,500; **Supplemental Figure 2, A-C**). The samples were then re-integrated using the
153 procedure described above. UMAP embeddings were calculated on the top 35 PCs with n=100 neighbors.
154 Epithelial compartments (basal, suprabasal, and superficial) were identified according to expression of
155 established markers (**Supplemental Figure 4D**)[27,28,47-50]. For each compartment, we computed composite
156 gene signature scores using Seurat's AddModuleScore() function (**Supplemental Figure 2E**). Cells were
157 initially clustered on an SNN graph with k=50 with resolution=0.25 (**Supplemental Figure 2F**), but
158 because the modularity optimization clustering implemented in Seurat grouped proliferating cells across
159 different epithelial compartments and inconsistently distinguished suprabasal and superficial cells, we
160 applied K-means clustering on the expression of the compartment-specific markers to delineate epithelial
161 compartments. We performed separate K=2 K-means clustering on 1) the Seurat clusters containing basal
162 and early suprabasal cells (clusters 1-5) to distinguish basal from suprabasal compartments and 2) the Seurat
163 clusters containing late suprabasal and superficial cells (clusters 6-8) to demarcate suprabasal and
164 superficial compartments (**Supplemental Figure 2, G-H**). Proliferating epithelial cells were identified
165 using Seurat's CellCycleScoring() function. Cells predicted as being in S or G2/M phases based on the
166 relative expression of corresponding markers[51] were labeled as proliferating cells (**Supplemental Figure**
167 **2I**). We tested for statistical differences in proportions of epithelial cell populations between conditions
168 using MASC[52]. To derive a continuous epithelial differentiation score for each cell (**Supplemental Figure**
169 **3**), we combined the compartment-specific module scores as follows :
170 $$EDS_C = score_{suprabasal} + score_{superficial} - score_{basal}$$
171

172 Where EDS_C is the epithelial differentiation score for a given cell. This EDS_C was then scaled between 0
173 and 1:
174

175 $$EDS_{C,scaled} = \frac{EDS_C - \min(EDS_1, \dots, EDS_n)}{\max(EDS_1, \dots, EDS_n) - \min(EDS_1, \dots, EDS_n)}$$

176

177 **Gene expression analysis**

178 To identify dysregulated pathways in SSc, we first calculated differential gene expression between
179 conditions within each epithelial compartment, considering each cell as a sample. Differential gene
180 expression was calculated using the FindMarkers() function in Seurat with the Wilcoxon rank-sum test on
181 log-normalized RNA counts for genes expressed in at least 1% of cells by compartment. We adjusted for
182 multiple testing using a Bonferroni correction accounting for the number of genes tested. Genes with an
183 adjusted $P<0.05$ and absolute $\log_2$ fold change >0.1 were considered significantly differentially expressed.
184 [53,54] Genes were then ranked by the magnitude and statistical certainty of their expression differences:
185

186 $$Weight = -\log_{10}(p) * abs(\log_2[FC])$$

187

188 Gene set enrichment was calculated using 3,795 canonical pathway gene sets from the Human Molecular
189 Signatures Database[55] using gene set enrichment analysis (GSEA)[53,54]. Statistical gene set enrichment was
190 calculated using one-way positive enrichment with 100,000 permutations on pathways with at least 10
191 genes. Mitochondrial and ribosomal genes were excluded from the GSEA.

192 To more robustly characterize gene expression differences between conditions[56], we performed
193 pseudo-bulk differential expression analysis by aggregating RNA counts per sample within epithelial
194 compartments. Genes with at least one count per million detected in at least 75% of samples in at least one
195 condition per epithelial compartment were analyzed. To mitigate outlier-driven signals, genes with log-
196 transformed expression greater than 3 standard deviations from the mean in exactly one individual were

197 excluded. Differential gene expression was evaluated using edgeR v3.42.4[57-59]. Gene expression was
198 modeled using a quasi-likelihood negative binomial generalized linear model, with global negative-
199 binomial trended dispersion estimated across all samples and gene-specific quasi-likelihood dispersions
200 estimated within each epithelial compartment. Differences in gene expression across conditions were
201 assessed using quasi-likelihood F-tests with false-discovery rate (FDR) adjustment. We conducted tests
202 combining both proximal and distal esophageal regions, as well as separate tests for each esophageal region.
203 Transcription factor enrichment analyses were conducted on the pseudo-bulk results using the
204 enricher() function from clusterProfiler v4.8.3[60]. We used consensus results from ENCODE[61] and ChEA[62],
205 as curated by the Ma'ayan Lab[63] and available at http://amp.pharm.mssm.edu/Enrichr/[64]to test for the
206 enrichment of transcription factor target genes from among genes that were differentially expressed
207 between SSc, GERD, and HCs with Bonferroni-adjusted $p<0.05$. Background gene sets were comprised of
208 the genes included in the pseudo-bulk differential expression analyses for each epithelial compartment.
209 Transcription factors absent from the background gene set were removed as candidate transcription factors.
210 Separate enrichment tests were performed for all differentially expressed genes (DEGs) and those that were
211 upregulated or downregulated in SSc. Enrichment results were adjusted using a Bonferroni correction
212 accounting for the number of transcription factors with more than 2 genes in the significant DEG list .
213
214 **Pseudo-bulk sample permutation**
215 Because of the sample size imbalance by disease state, which inherently yields different statistical power
216 for different tests across conditions, we permuted differential expression testing with equivalent sample
217 sizes to better evaluate the relative extent of dysregulation by EEC compartment and biopsy location
218 between SSc and GERD. We permuted all possible combinations of SSc samples with sample size n=4 for
219 both the proximal and distal regions and used the median statistics in the SSc vs. HCs tests to more directly
220 compare the results with those observed in GERD vs. HCs.
221
222 **Gene expression trajectory analysis**

223 To model gene expression as a continuous function along the axis of epithelial cell differentiation, we
224 utilized the framework implemented in Lamian[65], with the epithelial differentiation score serving as the
225 underlying pseudotime metric. Lamian fits a polynomial B-spline curve with the number of knots
226 determined by minimization of the Bayesian Information Criterion. With few cells near the lower and upper
227 limits of our epithelial differentiation score and the interpolation range bound by the minimum and
228 maximum pseudotime values, the B-spline models produced erratic fits near the interpolation boundaries.
229 Therefore, we instead modeled natural cubic splines, which are constrained to linearity at the boundaries
230 (second derivative is zero), and we excluded the lowest and highest 1% of cells from the interpolation
231 interval. Using these modeled gene expression trajectories, we performed hierarchical clustering of
232 differential gene trajectories for all genes that were significantly differentially expressed between SSc and
233 HC in single-cell-level differential expression testing.
234

235 **Clinical phenotype associations**

236 To test for associations with categorical measures of esophageal function (FLIP, HRM, and a combined
237 FLIP and HRM with NeuroMyogenic [NM] classification[66]), we performed ordinal logistic regression that
238 modeled likelihoods of increasing clinical severity. We excluded individuals with achalasia or
239 esophagogastric junction outflow obstruction (EGJOO) from ordinal regression modeling motility. For
240 FLIP we combined the "borderline/diminished" and "impaired/disordered" phenotypes. For NM we
241 combined the Stage I and Stage II ineffective motility phenotypes. The resultant ordering for the ordinal
242 regression was "normal"< "ineffective" < "absent" for HRM; "normal" < "borderline/diminished &
243 impaired/disordered" < "absent" for FLIP; and "normal" < "stages I & II – ineffective" < "stage III –
244 absent" for NM.

245 For association testing with quantitative clinical measures in SSc, we first performed a principal
246 components analysis (PCA) on a set of nine quantitative clinical traits: HRM mean distal contractile interval
247 (DCI), HRM basal esophagogastric junction (EGJ) pressure, HRM EGJ contractile index, HRM median
248 integrated relaxation pressure (IRP), FLIP intra-balloon pressure at 60ml, FLIP EGJ distensibility index,

249 the GerdQ impact score, the BEDQ score, and the NEQOL score. Missing values were imputed with the
250 mean. We then modeled associations with the first two PCs using linear regression. Gene expression was
251 represented using log-transformed counts.
252
253
254 **RESULTS**
255 Whole tissue single-cell RNA sequencing (scRNA-seq) was performed on paired proximal and distal
256 mucosal biopsies from ten patients with systemic sclerosis (SSc), four with gastroesophageal reflux disease
257 (GERD), and six healthy controls (HCs). All participants but one in each group were female (**Table 1**). SSc
258 patients were significantly older than HCs (54.9 vs 27.0 years; P<0.001). Among SSc patients, six had the
259 limited cutaneous subtype, three had the diffuse cutaneous subtype, and one had sine scleroderma with no
260 cutaneous symptoms (**Supplemental Table 1**). Mean SSc disease duration at the time of biopsy was 132
261 months. All SSc patients had positive serum antinuclear antibody and impaired esophageal motility.
262 Following sample-level quality control (**Supplemental Table 2**), 39 samples were retained for analysis,
263 including 19 from the proximal esophagus and 20 from the distal esophagus.
264
265 **Characterization of esophageal mucosal cell populations**
266 A total of 306,372 esophageal cells (7,856 mean cells per sample) were retained for cell clustering and gene
267 expression analysis (**Figure 1**). Each cell type was identified according to the expression of established
268 cell-specific gene markers including *KRT6A, KRT13,* and *KRT15* for epithelial cells (**Supplemental Table**
269 **3**, **Supplemental Figure 1**). Epithelial cells comprised the bulk of the sample (n=264,858; 86.45%),
270 followed by lymphoid cells (n=25,129; 8.20%), myeloid cells (n=10,019; 3.27%), endothelial cells
271 (n=4,027; 1.31%), and all other cell types (n=2,339; 0.76%).
272
273 **Significant loss of superficial EECs in SSc**

274 Following isolation of the epithelial cell cluster and additional quality control (**Supplemental Figure 2, A-**
275 **C**), we characterized the remaining 230,720 esophageal epithelial cells (EECs) according to their
276 differentiation and cell cycling states (**Figure 2A**). We classified the EECs into five primary compartments:
277 basal (n=55,818; 24%), proliferating basal (n=36,439; 16%), proliferating suprabasal (n=25,314; 11%),
278 suprabasal (n=82,283; 36%), and superficial (n=30,866; 13%) based on relative expression of canonical
279 epithelial genes and cell cycle markers (**Figure 2, B-C; Supplemental Figure 2, D-E**). Roughly 40% of
280 basal cells and 24% of suprabasal cells were proliferating, and the fractions of cells that were proliferating
281 did not differ by disease (**Figure 2D; Supplemental Figure 2, F-I**). Proportions of basal and suprabasal
282 cells were not significantly different across conditions, but there were significantly fewer superficial cells
283 in SSc compared to HCs (μ_{SSc}=0.10, μ_{HC}=0.21, P=0.003, **Figure 2E**). The reduction of superficial cells in
284 SSc was further apparent when examining EEC differentiation score distributions (**Supplemental Figure**
285 **3**) and projecting UMAP embeddings separately by disease state (**Figure 3**).
286
287 **EEC gene dysregulation highly correlated in SSc and GERD, limited to superficial cells**
288 Overall, the gene expression correlation between biopsy locations was extremely high, ranging from r=0.98
289 to r=0.99 for the 2,000 most variable genes (**Supplemental Figure 4A**) aggregated by condition and
290 epithelial compartment (**Supplemental Figure 4B**; **Supplemental Table 4**). Inter-sample expression
291 correlations indicated that the greatest gene expression heterogeneity was in the superficial compartment
292 (mean=0.94, std. dev.=0.027), followed by the basal compartment (mean=0.94, std. dev.=0.018; **Figure 4,**
293 **A-B**). Differential gene expression between conditions was most predominant in the superficial
294 compartment (**Figure 4, C-E**). At the single-cell level, about 6.6% and 4.1% of expressed genes were
295 significantly differentially expressed between SSc and HCs in the superficial compartment in the proximal
296 and distal regions, respectively, compared to just 1.3-1.4% and 1.3-3.0% in non-proliferating basal and
297 suprabasal cells (**Supplemental Tables 5-6**). Most differentially expressed genes (DEGs) were only
298 differentially expressed in one epithelial compartment (**Figure 4, C-D**). We observed more suprabasal
299 differential expression in the distal esophagus than in the proximal esophagus for both SSc and GERD

300 (**Figure 4, C-E**). Many of the same genes were significantly dysregulated in both SSc and GERD compared
301 to HCs (**Supplemental Table 5**). Depending on the compartment and biopsy location, 22-44% of
302 significantly dysregulated genes in SSc or GERD were differentially expressed in both conditions (**Figure**
303 **4E**). Consequently, the DEGs in SSc and GERD were significantly enriched for some of the same pathways,
304 particularly those related to the extracellular matrix and keratinization (**Figure 4F**). Clustered trajectories
305 of relative gene expression (SSc vs. HCs) by epithelial differentiation score for all single-cell-level DEGs
306 revealed distinct patterns of differential expression (**Figure 4, G-H**), which were principally distinguished
307 by the direction of gene expression change in superficial cells. Gene expression changes were more
308 pronounced in the most terminally differentiated superficial cells in the proximal esophagus (**Figure 4G**),
309 whereas in the distal esophagus the differences appeared more in late-stage suprabasal and early-stage
310 superficial cells (**Figure 4H**).

311 Six pathways were significantly enriched among DEGs between SSc and HCs: matrisome
312 (suprabasal, distal $P_{FDR}=2.3\times10^{-4}$, **Figure 4H**; superficial, distal $P_{FDR}=2.6\times10^{-2}$) and matrisome-associated
313 genes (suprabasal, distal $P_{FDR}=4.4\times10^{-4}$); cornified envelope formation (suprabasal, distal $P_{FDR}=2.1\times10^{-3}$;
314 superficial, distal $P_{FDR}=4.7\times10^{-2}$), and keratinization genes (suprabasal, distal $P_{FDR}=2.1\times10^{-3}$; superficial,
315 distal $P_{FDR}=5.0\times10^{-2}$); developmental biology-related genes (suprabasal, distal $P_{FDR}=4.1\times10^{-4}$; superficial,
316 distal $P_{FDR}=9.2\times10^{-3}$); and innate immune system genes ($P_{FDR}=2.8\times10^{-2}$, **Figure 4G**). No pathways were
317 significantly enriched among genes differentially expressed between SSc and GERD in any epithelial
318 compartment. The leading edges of significantly enriched pathways (**Supplemental Table 7**) contained
319 recurring genes from shared protein families, including serine protease inhibitors (serpins, **Supplemental**
320 **Figure 4C**), keratins (**Supplemental Figure 4D**), S100 proteins (**Supplemental Figure 4E**), and small
321 proline-rich proteins (SPRRs, **Supplemental Figure 4F**). The most frequent gene in leading edges of
322 significantly enriched pathways was *PI3* (**Supplemental Figure 4G**), including all pathways enriched in
323 SSc. Among the other most enriched pathways specific to SSc were those related to the regulation of metal
324 and immune homeostasis (**Figure 4F**, **Supplemental Table 7**).

325 Upon aggregating single-cell expression counts by sample for each EEC compartment, we found
326 that gene expression patterns clustered by EEC compartment (**Supplemental Figure 5A**) and that
327 differential expression between conditions was nearly exclusively limited to superficial cells (**Figure 5, A-**
328 **B; Supplemental Figure 5B**). There were 3,572 genes differentially expressed between SSc and HCs with
329 FDR q<0.05 in the superficial compartment, compared to just 232 total in all other compartments, including
330 only 172 that were not also differentially expressed in the superficial compartment (**Supplemental Table**
331 **8**). The differential expression observed in SSc was again highly correlated with the changes seen in GERD
332 (**Figure 5, C-D**). The correlations in $\log_2$ fold change between SSc and GERD vs. HCs were consistently
333 between 0.40-0.53 in all compartments except for superficial cells, where we observed much stronger
334 correlation in the proximal region (r=0.82) than in the distal region (r=0.36; **Supplemental Figure 5C**).
335 We detected 2,103 genes that were significantly differentially expressed in SSc but not GERD in
336 the proximal esophagus and 282 in the distal esophagus, of which 170 were differentially expressed only
337 in the distal region (**Figure 5, C-D**; **Supplemental Table 9**). Among the most upregulated SSc-specific
338 genes (**Figure 5, C-E; Supplemental Figure 5, C-D**) were genes related to inflammation (*PTGES*,
339 *MFGE8*)[67,68], innate immune response (*FCGBP, BST2*, *CD44*, *APOBEC3A*)[69-72], immune cell migration
340 (*C10orf99*, *LTB4R*, *ACKR3*)[73-75], antigen presentation (*HLA-B*, *CD74*, *TAP1*, *PSMB8*, *PSMB9*)[76], natural
341 killer cell activation (*CLEC2B*)[77], and fibroproliferation (*SGK1*, *HBEGF*)[78,79]. Four of the top five and eight
342 of the top 20 most downregulated SSc-specific genes by fold change in superficial EECs in the proximal
343 esophagus were metallothioneins (*MT1A*, *MT1E*, *MT1F*, *MT1G*, *MT1H*, *MT1M*, *MT1X*, *MT2A*). The
344 metallothioneins comprised the bulk of three pathways uniquely enriched in SSc with P<0.001: zinc
345 homeostasis, copper homeostasis, and cellular responses to stimuli (**Figure 4F**). The most down-regulated
346 SSc-specific gene across both the proximal and distal esophagus was the *H19* lncRNA. Interestingly, the
347 most down-regulated gene by fold change in GERD, *MUC22*, was not differentially expressed in SSc. One
348 gene, *SLC8A1-AS1*, was significantly differentially expressed in both GERD and SSc, but in opposite
349 directions (**Figure 5, C-E; Supplemental Table 9**).

350 More genes were significantly differentially expressed in SSc than in GERD, but in terms of
351 relative fold change vs. HCs, we observed greater expression changes in GERD (**Supplemental Figure**
352 **5C**). We therefore performed all possible sample permutations with equal numbers of SSc and GERD
353 samples (n=4) and repeated differential gene expression testing to more directly compare the relative
354 number of DEGs in both conditions (**Figure 5F**). In superficial cells from the proximal esophagus, there
355 were more DEGs in SSc in 60% of permutations, but the relative FC differences were greater in GERD in
356 61% of permutations (**Supplemental Table 10**). In the distal region, there were more DEGs in GERD in
357 62% of permutations, and the relative FC differences were greater in GERD in all permutations. The
358 correlations between SSc and GERD in terms of relative gene expression changes were consistently much
359 higher in the proximal region permutations than in the distal region (**Figure 5, G-H**). The permutation
360 results confirmed that the differences in gene expression observed in SSc relative to GERD were not simply
361 due to differences in sample size.
362 In summary, the gene expression changes measured in SSc and GERD were highly correlated. In
363 the distal esophagus, differential expression was more pronounced in GERD than in SSc, whereas in the
364 proximal esophagus there were more DEGs in SSc. In both conditions, differential gene expression was
365 most prevalent in the superficial compartment, but the expression changes appeared comparatively earlier
366 in the differentiation of distal EECs relative to proximal EECs.
367
368 **Transcription factor enrichment analysis**
369 We next examined whether DEGs in SSc superficial cells were significantly enriched for targets of specific
370 transcription factors. Among genes differentially expressed in SSc compared to GERD, the target genes of
371 IRF1 were significantly enriched (**Table 2**). IRF1 targets were significantly enriched among all proximal
372 superficial DEGs (p_{adj}=0.021) and among only those that were downregulated (p_{adj}=0.0024). However,
373 *IRF1* was not itself differentially expressed in superficial cells in SSc compared to GERD. Compared to
374 HCs, three transcription factors were significantly enriched in SSc in the proximal esophagus, including
375 MYC ($p_{adj}=2.1\times10^{-7}$) and E2F4 ($p_{adj}=5.5\times10^{-5}$) for upregulated DEGs and NFE2L2 for downregulated

376 DEGs (p_{adj}=7.6×10^{-4}). *MYC* and *E2F4* were also themselves differentially expressed in SSc compared to
377 HCs in superficial cells (p_{adj}=7.8×10^{-5}, p_{adj}=9.8×10^{-4}, respectively). MYC was the most enriched
378 transcription factor in genes upregulated in GERD compared to HCs, but the enrichment was not significant
379 after adjusting for multiple testing (p_{adj}=0.11). No transcription factors were significantly enriched in any
380 pairwise comparison in the distal esophagus.
381
382 **Expression mapping in superficial EECs**
383 Examining the expression patterns within superficial cells from the proximal esophagus more closely, we
384 identified five distinct clusters (**Figure 6A**) distinguished by unique expression marker patterns (**Figure**
385 **6B, Supplemental Figure 6, Supplemental Table 11**) and varying degrees of differentiation **(Figure 6C)**.
386 Clusters 1 and 3 represent the outermost, terminally differentiated epithelial layers, as evidenced by their
387 *FLG* expression[80-82]. They differed by the expression of metallothioneins, which were predominantly
388 expressed in cluster 3 (**Figure 6, E-F**). Cluster 3 was less abundant in SSc compared to HC with P<0.05
389 (**Figure 6D**), but metallothionein expression was lower in other superficial clusters, as well (**Figure 6F**).
390 Within the superficial EECs, the relative expression of metallothioneins was significantly correlated with
391 the differentiation score (r=0.06, P=2.3×10^{-23}). Cluster 2 was distinguished by its relatively high *CTSV*
392 expression. Clusters 4 and 5 were the least differentiated of the superficial cells and were distinguished by
393 their relative expression of *GJB6* and *FGFBP1*, respectively, although neither gene was differentially
394 expressed between conditions.
395 For the expression of the significantly enriched transcription factors (IRF1, MYC, E2F4, and
396 NFE2L2; **Table 2**) and their targets, we observed two consistent patterns (**Figure 7**). First, the greatest
397 magnitude differences in median target expression between SSc and HCs were seen in cluster 5. Second,
398 in each of those instances, the median target expression in GERD was between that of SSc and HCs. The
399 expression correlations between the enriched transcription factors and their targets were positive for *MYC*
400 (r=0.42) and *E2F4* (r=0.24) and negative for *IRF1* (r=-0.20) and *NFE2L2* (r=-0.18).

401

402 ***FLI1* not expressed in human EECs**

403 Deletion of the *Fli1* gene in epithelial cells in mice has been shown to recapitulate histological and
404 molecular features of esophageal involvement in SSc[24]. Therefore, we investigated the expression of *FLI1*
405 in our human esophageal samples. The expression of *FLI1* in human EECs was negligible and did not meet
406 the minimum expression thresholds for inclusion in our differential gene expression analysis. We did
407 observe low *FLI1* expression levels in endothelial cells (**Supplemental Figure 7A**), but the expression
408 differences between SSc, HCs, and GERD was not significant (**Supplemental Figure 7B**). There were
409 many FLI1 downstream targets among the proximal, superficial DEGs between SSc and HCs and seven
410 between SSc and GERD (*ARPC2*, *RAB24*, *MTPN*, *PPIF* upregulated; *CEBPZ*, *TRMT10C*, *TBP*
411 downregulated); however, the proportion of these genes among all DEGs was not different than what would
412 be expected by chance (P>0.05), and the average expression of FLI1 targets was higher in GERD and lower
413 in HCs compared to the expression in SSc (**Supplemental Figure 7C**).

414

415 **Correlations with clinical phenotypes**

416 We next evaluated whether the cellular and molecular changes we observed between SSc, GERD, and HCs
417 were correlated with specific clinical measures of esophageal dysfunction (**Table 1**). To test for associations
418 with esophageal motility phenotypes, we performed ordinal regression against increasing phenotype
419 severity. To more efficiently model the covariance structure among quantitative, functional esophageal
420 measures in SSc (**Figure 8A**), we performed PCA on a set of nine clinical metrics and tested for correlations
421 with the first two PCs (**Figure 8B**). The first PC explained 45.5% of variance and was most correlated with
422 FLIP intra-balloon pressure at 60ml (r=0.94), The second PC explained 22.6% of variance and was most
423 correlated with HRM basal EGJ pressure (r=0.74).

424 We did not observe any significant correlations between epithelial cell compartment proportions in
425 SSc and esophageal motility phenotypes. It appeared that the low proportion of superficial cells in the
426 esophageal epithelium was a universal feature in SSc, regardless of phenotype (**Supplemental Figure 8A**).

427 The proportion of superficial cells decreased with disease duration, but the trend was not statistically
428 significant (P=0.1; **Supplemental Figure 8B**). We also did not observe any significant correlations between
429 EEC compartment proportions and quantitative trait PCs (**Figure 8C**).

430 We then tested for correlations with gene expression in superficial EECs for genes that were
431 significantly differentially expressed in SSc compared to HCs (P_{adj}<0.05) and nominally differentially
432 expressed in SSc compared to GERD (P<0.05). There were 433 genes that met these criteria in the proximal
433 esophagus and 99 in the distal esophagus. Although no associations with esophageal motility phenotypes
434 were statistically significant after adjusting for multiple testing, the strongest phenotypic correlation was
435 with *TRIM11* (P_{FLIP}=0.01, P_{HRM}=0.007, **Supplemental Figure 8C**), a gene recently found to attenuate Treg
436 cell differentiation in $CD4^+$ T cells in mice[83]. No associations with the top two clinical PCs were significant
437 after adjusting for multiple testing. We also tested for clinical PC correlations with aggregate
438 metallothionein gene expression and transcription factor target expression for IRF1, MYC, E2F4, and
439 NFE2L2. We observed one nominal association between the mean metallothionein module score and PC1
440 (P=0.02, **Figure 8D**). Notably, within superficial cells we also observed a strong correlation between the
441 proportion of metallothionein-expressing cluster 3 cells and PC1 in the distal esophagus (**Figure 8E**). This
442 indicates that not only is this population of cells reduced in SSc, but its relative decrease is further correlated
443 with more severe esophageal involvement (**Figure 8F**). Therefore, our findings suggest that increased
444 esophageal dysmotility in SSc is associated with a decline in metallothionein-expressing superficial EECs,
445 which coincides with lower overall metallothionein expression in superficial cells.

446

447

448 **DISCUSSION**

449 Esophageal dysfunction is extremely common in SSc and is significantly associated with increased
450 mortality and lower quality of life[84,85]. The causal mechanisms by which SSc affects the esophagus,
451 however, have not been conclusively determined. Here, we sought to clarify the role of epithelial cells in
452 SSc esophageal dysfunction by using scRNA-seq to quantify cellular and transcriptional changes relative

453 to HCs and individuals with GERD. While our findings indicate that epithelial changes in SSc result
454 primarily from chronic acid exposure, they also highlight immunoregulatory pathways uniquely altered in
455 SSc that may be linked to pathogenic aberrations.

456 With epithelial layers of the skin and internal organs being primary sites of injury in SSc, the
457 pathogenic role of epithelial cells in SSc has long been an ongoing area of research. When damaged,
458 epithelial cells release signals that help induce fibroblast activation to promote wound healing[86]. In SSc,
459 epidermal keratinocyte characteristics resemble a pro-fibrotic, activated state[87]. SSc epidermal cells were
460 also found to stimulate fibroblasts in culture[88], and some signs related to epithelial-mesenchymal transition
461 (EMT) have been observed in the SSc epidermis[89]. The FLI1 (friend leukemia integration 1) transcription
462 factor, in particular, has received much attention for its potential role in epithelial-cell-mediated SSc
463 pathogenesis[20]. Lower FLI1 expression was observed in the epidermis of diffuse cutaneous SSc, and
464 inactivation of FLI1 in human keratinocytes in vitro induced gene expression changes characteristic of
465 SSc[24]. Furthermore, conditional deletion of *Fli1* in epithelial cells produced an SSc-like phenotype in
466 mice[24]. Importantly, the epithelial-cell-specific *Fli1* knockout further recapitulated the esophageal
467 involvement of SSc, with increased collagen deposition in the lamina propria combined with atrophy of the
468 circular muscle layer, and the esophageal changes were not mediated by T-cell autoimmunity[24]. In contrast
469 to these previous studies in mice and epidermal keratinocytes, we did not observe *FLI1* expression in EECs
470 in any condition. *FLI1* was expressed at low levels in endothelial cells, but the expression differences
471 between SSc, HCs, and GERD were not significant. We did observe significantly increased expression in
472 SSc and GERD of the *PI3* gene, which encodes trappin-2/elafin, previously found to be induced by Fli1
473 silencing in human dermal microvascular endothelial cells[90]. However, this association is unlikely to be
474 pathogenic, as the increase in gene expression was greater in GERD, and a growing body of evidence
475 indicates that trappin-2/elafin is expressed to promote tissue repair in response to gastrointestinal tissue
476 inflammation[91].

477 While there have been many studies of epithelial cells in skin in SSc, molecular study of the human
478 esophagus in SSc had heretofore been limited to one array-based gene expression study in bulk tissue

479 conducted by Taroni and colleagues[15], in which they identified distinct expression signatures among
480 variable genes between 15 individuals with SSc that were similar to those seen in skin[92]. These signatures
481 were independent of clinically defined SSc subsets[15], suggestive of inter-individual heterogeneity in SSc.
482 They also found that gene expression patterns between upper and lower esophageal biopsies within
483 individuals were nearly identical. They did not observe statistical associations between inflammatory gene
484 expression signatures and GERD, but the study was underpowered for such an analysis, and GERD was
485 defined based on histological evidence for basal cell hyperplasia and intraepithelial lymphocyte counts.

486 The gene expression differences we observed in SSc epithelial cells were highly correlated with
487 those seen in GERD, and nearly all differential gene expression was limited to the superficial layers of the
488 epithelium. These observations suggest that the primary driver of differential gene expression in EECs in
489 SSc is chronic acid exposure. The genes with the strongest mutual upregulation in SSc and GERD were
490 primarily related to keratinization/cornification, including small proline-rich proteins (SPRRs), serine
491 protease inhibitors (serpins), S100 proteins, late cornified envelope (LCE) genes, and keratins associated
492 with terminal epithelial differentiation, namely *KRTDAP* and *KRT*. *KRT1* helps maintain epithelial barrier
493 function in gastrointestinal epithelial cells, and its overexpression was shown to attenuate IL-1β-induced
494 epithelial permeability[93]. The mutual upregulation of these genes in the outer layers of the epithelium
495 therefore likely represent a protective response to chronic acid exposure.

496 While the overall gene expression differences were highly correlated between SSc and GERD,
497 there were sets of genes disproportionately upregulated in SSc that hint at disease-related immune response
498 aberrations, including genes that have previously been implicated in SSc pathogenesis. For example,
499 *PTGES*, which encodes prostaglandin E synthase, was previously found to be significantly upregulated in
500 SSc fibroblasts[94] and inflammatory non-classical monocytes[95], and Ptges-null mice were resistant to
501 bleomycin-induced fibrosis[94]. CD44 and CD74 form the receptor complex for the macrophage inhibitory
502 factor (MIF), an inflammatory cytokine that promotes fibroblast migration and has been implicated in SSc
503 pathogenesis[96]. LTB4R (Leukotriene B4 Receptor, or BLT1) has been found to activate AKT/mTOR
504 signaling and knockdown of *LTB4R* attenuated fibrosis in bleomycin-induced SSc in murine models[97].

505 Serum HBEGF (heparin-binding epidermal growth factor) levels and fibroblast *HBEGF* expression were
506 both significantly elevated in SSc[98]. Copy number variation of *APOBEC3A* was significantly associated
507 with SSc in a Han Chinese population[99]. Finally, the SSc-specific upregulation of genes involved in antigen
508 processing and presentation (*HLA-B*, *CD74*, *TAP1*, *PSMB8*, *PSMB9*) are suggestive of increased human
509 leukocyte antigen (HLA) class I activity in superficial EECs in SSc. Notably, *TAP1*, *PSMB8*, and *PSMB9*
510 are located immediately adjacent to one another in the class II region of the HLA locus, which is the
511 genomic region with the strongest genetic associations with SSc[100], and SSc-associated HLA-B alleles have
512 also been identified[101,102].

513 The most striking set of genes that were uniquely downregulated in SSc were metallothioneins.
514 The expression of detected metallothioneins (*MT1A*, *MT1E*, *MT1F*, *MT1G*, *MT1H*, *MT1M*, *MT1X*, *MT2A*)
515 was strongly induced in the superficial compartment of EECs, but significantly less so in SSc and,
516 interestingly, only in the proximal esophagus. We further observed a nominal association between relative
517 metallothionein expression in the proximal esophagus and the first PC of esophageal dysmotility among
518 SSc patients, with lower expression of metallothioneins correlated with greater EGJ distensibility and
519 weaker contractility. The proportion of the primary metallothionein-expressing cluster within superficial
520 EECs was likewise significantly associated with clinical PC1. The metallothioneins are a family of well-
521 conserved, metal-binding proteins that regulate zinc and copper homeostasis, prevent heavy metal
522 poisoning, and combat oxidative stress[103,104]. Their study in a wide array of physiological conditions and
523 immune-mediated diseases has revealed the proteins play central roles in innate and adaptive immunity,
524 including autoimmunity, but their immunoregulatory behavior is highly complex and context-specific[103,104].
525 Given their import in immune regulation, their associations with SSc and esophageal motility reported in
526 this study, and the connection between SSc and heavy metal exposure[105], the metallothioneins are
527 compelling candidates that warrant further study in SSc.

528 Interestingly, there was greater relative gene dysregulation in the proximal esophagus compared to
529 the distal esophagus in SSc, but the opposite was true for GERD. This discrepancy suggests that esophageal
530 dysmotility in SSc leads to greater acid exposure in the proximal esophagus, since refluxed gastric acid

531 typically affects the distal esophagus significantly more than the proximal esophagus[106] and abnormal
532 peristalsis prolongs acid clearance[107]. This finding coincides with a bulk RNA-seq study of esophageal
533 mucosa of achalasia patients[108], which likewise identified more differential expression in the proximal
534 esophagus than in the distal esophagus, including significant enrichment of matrisome-associated genes
535 and lower expression of genes associated with reactive oxygen species metabolism[108]. Due to differences
536 in innervation depth[109], the epithelium of the proximal esophagus has more nociceptive sensitivity than the
537 distal esophagus, which likely contributes to the association between gastrointestinal symptoms and lower
538 quality of life in SSc[110]. Proximal acid exposure increases the risk of aspiration[111] and is significantly more
539 prevalent in SSc patients with idiopathic lung fibrosis[112].

540 Targets for several transcription factors, including IRF1, MYC, E2F4, and NFE2L2 (or NRF2),
541 were significantly enriched among DEGs in superficial EECs from the proximal esophagus. IRF1 is
542 involved innate immune responses, but while IRF1 targets were enriched in DEGs between SSc and GERD,
543 the enrichment appeared to be driven by upregulation in GERD. IRF1 targets were not enriched in DEGs
544 between SSc and HCs, nor was *IRF1* significantly differentially expressed in SSc. MYC, E2F4, and
545 NFE2L2 all participate in different stages of the cell cycle[113-115], and expression of each was highest in
546 either the proliferating basal or suprabasal compartments, but these proteins may also play a role in terminal
547 differentiation of EECs, particularly under conditions of stress. In the least differentiated superficial cluster
548 (cluster 5), where the expression differences relative to HCs for each of these transcription factors and their
549 targets was most pronounced, the expression level changes in GERD were correlated with those observed
550 in SSc. Furthermore, Myc ablation in Krt14-expressing tissues in mice led to abnormal terminal
551 differentiation of epithelial cells[116], and a study of the esophageal epithelium in Nrf2 knockout mice
552 indicated that Nrf2 regulates cornification of keratinized epithelial cells under conditions of stress[117].

553 There are important limitations to consider when interpreting this study. First, this study only
554 focused on squamous epithelial cells. There may be relevant changes in other cell types within the
555 esophageal mucosa that correlate with SSc and dysmotility, but such analyses are ongoing and were outside
556 the scope of this investigation. Second, the relatively small patient sample size of this study combined with

557 the heterogeneity of SSc subtypes and dysmotility phenotypes limited our power to detect clinical
558 associations between cell type proportions or gene expression changes and specific clinical phenotypes.
559 Heterogeneous molecular profiles in SSc esophagus samples have been described previously[15], but we
560 grouped all SSc patients together in our analyses for comparisons against GERD and HCs. We applied
561 ordinal logistic regression on motility classifications and performed principal components analysis on
562 reflux- and motility-related quantitative traits to increase our power to detect clinical associations but were
563 nonetheless statistically limited by sample size. Third, we did not examine imaging data in the SSc
564 esophageal mucosa. Therefore, we cannot make conclusions on the morphological states that accompany
565 the transcriptional and cellular proportion changes that we observed. Future studies utilizing larger, more
566 homogeneous cohorts with imaging data may be able to detect more esophageal epithelial changes
567 associated with dysmotility in SSc.

568 In summary, esophageal complications are common in individuals with SSc and can greatly impact
569 quality of life and lifespan. In this study, we sought to clarify the role of the epithelium in SSc esophageal
570 involvement. Through a thorough, single-cell-level transcriptomic investigation, we identified the unique
571 cellular and transcriptional differences present in the SSc esophageal epithelium. There were significantly
572 fewer terminally differentiated epithelial cells in the outermost, superficial layers in both the proximal and
573 distal esophagus, but otherwise epithelial compartment proportions were similar to healthy controls,
574 including proliferating cell proportions. Significant differences in gene expression were likewise almost
575 exclusively found in the superficial compartment, and these changes were highly correlated with those seen
576 in individuals with GERD, indicating that the primary driver of differential gene expression in the SSc
577 esophageal epithelium is chronic acid exposure. Transcriptional differences were also more prominent in
578 the proximal esophagus of SSc patients, suggesting that esophageal dysmotility leads to greater proximal
579 exposure. Genes that were most disproportionately dysregulated in SSc compared to GERD belonged to
580 immune-related pathways, including innate and adaptive response, antigen presentation, and metal
581 homeostasis, and may therefore point to pathogenic immune dysfunction. Future studies investigating these
582 pathways in non-epithelial cell populations in more homogeneous SSc cohorts may further resolve the

583 pathogenic mechanisms driving esophageal involvement in SSc. By serving as an atlas for the human
584 esophageal epithelium in SSc, this study can guide future efforts to address remaining gaps in our
585 understanding of the SSc esophagus, to ultimately uncover pathogenic mechanisms and identify actionable
586 targets.

587 **REFERENCES**

588 1. Bairkdar, M., Rossides, M., Westerlind, H., Hesselstrand, R., Arkema, E.V., and Holmqvist, M.
589 (2021). Incidence and prevalence of systemic sclerosis globally: a comprehensive systematic
590 review and meta-analysis. Rheumatology (Oxford) *60*, 3121-3133. 10.1093/rheumatology/keab190.
591 2. Fan, Y., Bender, S., Shi, W., and Zoz, D. (2020). Incidence and prevalence of systemic sclerosis
592 and systemic sclerosis with interstitial lung disease in the United States. J Manag Care Spec Pharm
593 *26*, 1539-1547. 10.18553/jmcp.2020.20136.
594 3. Tian, J., Kang, S., Zhang, D., Huang, Y., Zhao, M., Gui, X., Yao, X., and Lu, Q. (2023). Global,
595 regional, and national incidence and prevalence of systemic sclerosis. Clin Immunol *248*, 109267.
596 10.1016/j.clim.2023.109267.
597 4. Pope, J.E., Quansah, K., Hassan, S., Seung, S.J., Flavin, J., and Kolb, M. (2021). Systemic Sclerosis
598 and Associated Interstitial Lung Disease in Ontario, Canada: An Examination of Prevalence and
599 Survival Over 10 Years. J Rheumatol *48*, 1427-1434. 10.3899/jrheum.201049.
600 5. Li, J.X. (2022). Secular Trends in Systemic Sclerosis Mortality in the United States from 1981 to
601 2020. Int J Environ Res Public Health *19*. 10.3390/ijerph192215088.
602 6. Rubio-Rivas, M., Royo, C., Simeon, C.P., Corbella, X., and Fonollosa, V. (2014). Mortality and
603 survival in systemic sclerosis: systematic review and meta-analysis. Semin Arthritis Rheum *44*,
604 208-219. 10.1016/j.semarthrit.2014.05.010.
605 7. Bairkdar, M., Chen, E.Y., Dickman, P.W., Hesselstrand, R., Westerlind, H., and Holmqvist, M.
606 (2023). Survival in Swedish patients with systemic sclerosis: a nationwide population-based
607 matched cohort study. Rheumatology (Oxford) *62*, 1170-1178. 10.1093/rheumatology/keac474.
608 8. Li, B., Yan, J., Pu, J., Tang, J., Xu, S., and Wang, X. (2021). Esophageal Dysfunction in Systemic
609 Sclerosis: An Update. Rheumatol Ther *8*, 1535-1549. 10.1007/s40744-021-00382-0.
610 9. Emmanuel, A. (2016). Current management of the gastrointestinal complications of systemic
611 sclerosis. Nat Rev Gastroenterol Hepatol *13*, 461-472. 10.1038/nrgastro.2016.99.
612 10. Ahuja, N.K., and Clarke, J.O. (2021). Scleroderma and the Esophagus. Gastroenterol Clin North
613 Am *50*, 905-918. 10.1016/j.gtc.2021.08.005.
614 11. Denaxas, K., Ladas, S.D., and Karamanolis, G.P. (2018). Evaluation and management of
615 esophageal manifestations in systemic sclerosis. Ann Gastroenterol *31*, 165-170.
616 10.20524/aog.2018.0228.
617 12. Foocharoen, C., Thinkhamrop, W., Chaichaya, N., Mahakkanukrauh, A., Suwannaroj, S., and
618 Thinkhamrop, B. (2022). Development and validation of machine learning for early mortality in
619 systemic sclerosis. Sci Rep *12*, 17178. 10.1038/s41598-022-22161-9.
620 13. Sjogren, R.W. (1994). Gastrointestinal motility disorders in scleroderma. Arthritis Rheum *37*, 1265-
621 1282. 10.1002/art.1780370902.
622 14. Roberts, C.G., Hummers, L.K., Ravich, W.J., Wigley, F.M., and Hutchins, G.M. (2006). A case-
623 control study of the pathology of oesophageal disease in systemic sclerosis (scleroderma). Gut *55*,
624 1697-1703. 10.1136/gut.2005.086074.
625 15. Taroni, J.N., Martyanov, V., Huang, C.C., Mahoney, J.M., Hirano, I., Shetuni, B., Yang, G.Y.,
626 Brenner, D., Jung, B., Wood, T.A., et al. (2015). Molecular characterization of systemic sclerosis
627 esophageal pathology identifies inflammatory and proliferative signatures. Arthritis Res Ther *17*,
628 194. 10.1186/s13075-015-0695-1.
629 16. Roman, S., Hot, A., Fabien, N., Cordier, J.F., Miossec, P., Ninet, J., Mion, F., and Reseau
630 Sclerodermie des Hospices Civils de, L. (2011). Esophageal dysmotility associated with systemic
631 sclerosis: a high-resolution manometry study. Dis Esophagus *24*, 299-304. 10.1111/j.1442-
632 2050.2010.01150.x.
633 17. Noviani, M., Chellamuthu, V.R., Albani, S., and Low, A.H.L. (2022). Toward Molecular
634 Stratification and Precision Medicine in Systemic Sclerosis. Front Med (Lausanne) *9*, 911977.
635 10.3389/fmed.2022.911977.

18. Carlson, D.A., Crowell, M.D., Kimmel, J.N., Patel, A., Gyawali, C.P., Hinchcliff, M., Griffing, W.L., Pandolfino, J.E., and Vela, M.F. (2016). 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*, 1502-1506. 10.1016/j.cgh.2016.03.039.
19. Kimmel, J.N., Carlson, D.A., Hinchcliff, M., Carns, M.A., Aren, K.A., Lee, J., and Pandolfino, J.E. (2016). The association between systemic sclerosis disease manifestations and esophageal high-resolution manometry parameters. Neurogastroenterol Motil *28*, 1157-1165. 10.1111/nmo.12813.
20. Asano, Y., Takahashi, T., and Saigusa, R. (2019). Systemic sclerosis: Is the epithelium a missing piece of the pathogenic puzzle? J Dermatol Sci *94*, 259-265. 10.1016/j.jdermsci.2019.04.007.
21. Durcan, C., Hossain, M., Chagnon, G., Peric, D., Karam, G., Bsiesy, L., and Girard, E. (2022). Experimental investigations of the human oesophagus: anisotropic properties of the embalmed mucosa-submucosa layer under large deformation. Biomech Model Mechanobiol *21*, 1685-1702. 10.1007/s10237-022-01613-1.
22. Kou, W., Pandolfino, J.E., Kahrilas, P.J., and Patankar, N.A. (2017). Simulation studies of the role of esophageal mucosa in bolus transport. Biomech Model Mechanobiol *16*, 1001-1009. 10.1007/s10237-016-0867-1.
23. Asano, Y., Bujor, A.M., and Trojanowska, M. (2010). The impact of Fli1 deficiency on the pathogenesis of systemic sclerosis. J Dermatol Sci *59*, 153-162. 10.1016/j.jdermsci.2010.06.008.
24. Takahashi, T., Asano, Y., Sugawara, K., Yamashita, T., Nakamura, K., Saigusa, R., Ichimura, Y., Toyama, T., Taniguchi, T., Akamata, K., et al. (2017). Epithelial Fli1 deficiency drives systemic autoimmunity and fibrosis: Possible roles in scleroderma. J Exp Med *214*, 1129-1151. 10.1084/jem.20160247.
25. Dinh, H.Q., Pan, F., Wang, G., Huang, Q.F., Olingy, C.E., Wu, Z.Y., Wang, S.H., Xu, X., Xu, X.E., He, J.Z., et al. (2021). Integrated single-cell transcriptome analysis reveals heterogeneity of esophageal squamous cell carcinoma microenvironment. Nat Commun *12*, 7335. 10.1038/s41467-021-27599-5.
26. Zhang, X., Peng, L., Luo, Y., Zhang, S., Pu, Y., Chen, Y., Guo, W., Yao, J., Shao, M., Fan, W., et al. (2021). Dissecting esophageal squamous-cell carcinoma ecosystem by single-cell transcriptomic analysis. Nat Commun *12*, 5291. 10.1038/s41467-021-25539-x.
27. Rochman, M., Wen, T., Kotliar, M., Dexheimer, P.J., Ben-Baruch Morgenstern, N., Caldwell, J.M., Lim, H.W., and Rothenberg, M.E. (2022). Single-cell RNA-Seq of human esophageal epithelium in homeostasis and allergic inflammation. JCI Insight *7*. 10.1172/jci.insight.159093.
28. Clevenger, M.H., Karami, A.L., Carlson, D.A., Kahrilas, P.J., Gonsalves, N., Pandolfino, J.E., Winter, D.R., Whelan, K.A., and Tetreault, M.P. (2023). Suprabasal cells retain progenitor cell identity programs in eosinophilic esophagitis-driven basal cell hyperplasia. JCI Insight *8*. 10.1172/jci.insight.171765.
29. Carlson, D.A., Prescott, J.E., Germond, E., Brenner, D., Carns, M., Correia, C.S., Tetreault, M.P., McMahan, Z.H., Hinchcliff, M., Kou, W., et al. (2022). Heterogeneity of primary and secondary peristalsis in systemic sclerosis: A new model of "scleroderma esophagus". Neurogastroenterol Motil *34*, e14284. 10.1111/nmo.14284.
30. Carlson, D.A., Gyawali, C.P., Khan, A., Yadlapati, R., Chen, J., Chokshi, R.V., Clarke, J.O., Garza, J.M., Jain, A.S., Katz, P., et al. (2021). Classifying Esophageal Motility by FLIP Panometry: A Study of 722 Subjects With Manometry. Am J Gastroenterol *116*, 2357-2366. 10.14309/ajg.0000000000001532.
31. Koop, A.H., Kahrilas, P.J., Schauer, J., Pandolfino, J.E., and Carlson, D.A. (2023). The impact of primary peristalsis, contractile reserve, and secondary peristalsis on esophageal clearance measured by timed barium esophagogram. Neurogastroenterol Motil, e14638. 10.1111/nmo.14638.
32. Yadlapati, R., Kahrilas, P.J., Fox, M.R., Bredenoord, A.J., Prakash Gyawali, C., Roman, S., Babaei, A., Mittal, R.K., Rommel, N., Savarino, E., et al. (2021). Esophageal motility disorders on high-resolution manometry: Chicago classification version 4.0((c)). Neurogastroenterol Motil *33*, e14058. 10.1111/nmo.14058.

33. Stern, E.K., Carlson, D.A., Falmagne, S., Hoffmann, A.D., Carns, M., Pandolfino, J.E., Hinchcliff, M., and Brenner, D.M. (2018). Abnormal esophageal acid exposure on high-dose proton pump inhibitor therapy is common in systemic sclerosis patients. Neurogastroenterol Motil *30*. 10.1111/nmo.13247.
34. Yadlapati, R., Kahrilas, P.J., Fox, M.R., Bredenoord, A.J., Prakash Gyawali, C., Roman, S., Babaei, A., Mittal, R.K., Rommel, N., Savarino, E., et al. (2021). Esophageal motility disorders on high-resolution manometry: Chicago classification version 4.0(©). Neurogastroenterol Motil *33*, e14058. 10.1111/nmo.14058.
35. Rooney, K.P., Baumann, A.J., Donnan, E., Kou, W., Triggs, J.R., Prescott, J., Decorrevont, A., Dorian, E., Kahrilas, P.J., Pandolfino, J.E., and Carlson, D.A. (2021). Esophagogastric Junction Opening Parameters Are Consistently Abnormal in Untreated Achalasia. Clin Gastroenterol Hepatol *19*, 1058-1060.e1051. 10.1016/j.cgh.2020.03.069.
36. Carlson, D.A., Baumann, A.J., Donnan, E.N., Krause, A., Kou, W., and Pandolfino, J.E. (2021). Evaluating esophageal motility beyond primary peristalsis: Assessing esophagogastric junction opening mechanics and secondary peristalsis in patients with normal manometry. Neurogastroenterol Motil *33*, e14116. 10.1111/nmo.14116.
37. Carlson, D.A., Kou, W., Lin, Z., Hinchcliff, M., Thakrar, A., Falmagne, S., Prescott, J., Dorian, E., Kahrilas, P.J., and Pandolfino, J.E. (2019). Normal Values of Esophageal Distensibility and Distension-Induced Contractility Measured by Functional Luminal Imaging Probe Panometry. Clin Gastroenterol Hepatol *17*, 674-681 e671. 10.1016/j.cgh.2018.07.042.
38. Jonasson, C., Wernersson, B., Hoff, D.A., and Hatlebakk, J.G. (2013). Validation of the GerdQ questionnaire for the diagnosis of gastro-oesophageal reflux disease. Aliment Pharmacol Ther *37*, 564-572. 10.1111/apt.12204.
39. Taft, T.H., Riehl, M., Sodikoff, J.B., Kahrilas, P.J., Keefer, L., Doerfler, B., and Pandolfino, J.E. (2016). Development and validation of the brief esophageal dysphagia questionnaire. Neurogastroenterol Motil *28*, 1854-1860. 10.1111/nmo.12889.
40. Bedell, A., Taft, T.H., Keefer, L., and Pandolfino, J. (2016). Development of the Northwestern Esophageal Quality of Life Scale: A Hybrid Measure for Use Across Esophageal Conditions. Am J Gastroenterol *111*, 493-499. 10.1038/ajg.2016.20.
41. Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W.M., 3rd, Zheng, S., Butler, A., Lee, M.J., Wilk, A.J., Darby, C., Zager, M., et al. (2021). Integrated analysis of multimodal single-cell data. Cell *184*, 3573-3587 e3529. 10.1016/j.cell.2021.04.048.
42. Hippen, A.A., Falco, M.M., Weber, L.M., Erkan, E.P., Zhang, K., Doherty, J.A., Vaharautio, A., Greene, C.S., and Hicks, S.C. (2021). miQC: An adaptive probabilistic framework for quality control of single-cell RNA-sequencing data. PLoS Comput Biol *17*, e1009290. 10.1371/journal.pcbi.1009290.
43. Germain, P.L., Lun, A., Garcia Meixide, C., Macnair, W., and Robinson, M.D. (2021). Doublet identification in single-cell sequencing data using scDblFinder. F1000Res *10*, 979. 10.12688/f1000research.73600.2.
44. Hafemeister, C., and Satija, R. (2019). Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. Genome Biol *20*, 296. 10.1186/s13059-019-1874-1.
45. Lause, J., Berens, P., and Kobak, D. (2021). Analytic Pearson residuals for normalization of single-cell RNA-seq UMI data. Genome Biol *22*, 258. 10.1186/s13059-021-02451-7.
46. McInnes, L., Healy, J., and Melville, J. (2018). UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. arXiv:1802.03426. 10.48550/arXiv.1802.03426.
47. Clevenger, M.H., Karami, A.L., Carlson, D.A., Kahrilas, P.J., Gonsalves, N., Pandolfino, J.E., Winter, D.R., Whelan, K.A., and Tetreault, M.P. (2023). Suprabasal cells retaining stem cell identity programs drive basal cell hyperplasia in eosinophilic esophagitis. bioRxiv. 10.1101/2023.04.20.537495.

48. Busslinger, G.A., Weusten, B.L.A., Bogte, A., Begthel, H., Brosens, L.A.A., and Clevers, H. (2021). Human gastrointestinal epithelia of the esophagus, stomach, and duodenum resolved at single-cell resolution. Cell Rep *34*, 108819. 10.1016/j.celrep.2021.108819.
49. Kabir, M.F., Karami, A.L., Cruz-Acuna, R., Klochkova, A., Saxena, R., Mu, A., Murray, M.G., Cruz, J., Fuller, A.D., Clevenger, M.H., et al. (2022). Single cell transcriptomic analysis reveals cellular diversity of murine esophageal epithelium. Nat Commun *13*, 2167. 10.1038/s41467-022-29747-x.
50. Okumura, T., Shimada, Y., Imamura, M., and Yasumoto, S. (2003). Neurotrophin receptor p75(NTR) characterizes human esophageal keratinocyte stem cells in vitro. Oncogene *22*, 4017-4026. 10.1038/sj.onc.1206525.
51. Tirosh, I., Izar, B., Prakadan, S.M., Wadsworth, M.H., 2nd, Treacy, D., Trombetta, J.J., Rotem, A., Rodman, C., Lian, C., Murphy, G., et al. (2016). Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. Science *352*, 189-196. 10.1126/science.aad0501.
52. Fonseka, C.Y., Rao, D.A., Teslovich, N.C., Korsunsky, I., Hannes, S.K., Slowikowski, K., Gurish, M.F., Donlin, L.T., Lederer, J.A., Weinblatt, M.E., et al. (2018). Mixed-effects association of single cells identifies an expanded effector CD4(+) T cell subset in rheumatoid arthritis. Sci Transl Med *10*. 10.1126/scitranslmed.aaq0305.
53. Mootha, V.K., Lindgren, C.M., Eriksson, K.F., Subramanian, A., Sihag, S., Lehar, J., Puigserver, P., Carlsson, E., Ridderstrale, M., Laurila, E., et al. (2003). PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. Nat Genet *34*, 267-273. 10.1038/ng1180.
54. Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., and Mesirov, J.P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A *102*, 15545-15550. 10.1073/pnas.0506580102.
55. Liberzon, A., Birger, C., Thorvaldsdottir, H., Ghandi, M., Mesirov, J.P., and Tamayo, P. (2015). The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst *1*, 417-425. 10.1016/j.cels.2015.12.004.
56. Squair, J.W., Gautier, M., Kathe, C., Anderson, M.A., James, N.D., Hutson, T.H., Hudelle, R., Qaiser, T., Matson, K.J.E., Barraud, Q., et al. (2021). Confronting false discoveries in single-cell differential expression. Nat Commun *12*, 5692. 10.1038/s41467-021-25960-2.
57. Chen, Y., Lun, A.T., and Smyth, G.K. (2016). From reads to genes to pathways: differential expression analysis of RNA-Seq experiments using Rsubread and the edgeR quasi-likelihood pipeline. F1000Res *5*, 1438. 10.12688/f1000research.8987.2.
58. McCarthy, D.J., Chen, Y., and Smyth, G.K. (2012). Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. Nucleic Acids Res *40*, 4288-4297. 10.1093/nar/gks042.
59. Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139-140. 10.1093/bioinformatics/btp616.
60. Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., Feng, T., Zhou, L., Tang, W., Zhan, L., et al. (2021). clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. Innovation (Camb) *2*, 100141. 10.1016/j.xinn.2021.100141.
61. Consortium, E.P. (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57-74. 10.1038/nature11247.
62. Lachmann, A., Xu, H., Krishnan, J., Berger, S.I., Mazloom, A.R., and Ma'ayan, A. (2010). ChEA: transcription factor regulation inferred from integrating genome-wide ChIP-X experiments. Bioinformatics *26*, 2438-2444. 10.1093/bioinformatics/btq466.
63. Kou, Y., Chen, E.Y., Clark, N.R., Duan, Q., Tan, C.M., and Ma‘ayan, A. (2013). ChEA2: Gene-Set Libraries from ChIP-X Experiments to Decode the Transcription Regulome. held in Berlin, Heidelberg, (Springer Berlin Heidelberg), pp. 416-430.

788 64. Chen, E.Y., Tan, C.M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G.V., Clark, N.R., and Ma'ayan, A.
789 (2013). Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC
790 Bioinformatics *14*, 128. 10.1186/1471-2105-14-128.
791 65. Hou, W., Ji, Z., Chen, Z., Wherry, E.J., Hicks, S.C., and Ji, H. (2023). A statistical framework for
792 differential pseudotime analysis with multiple single-cell RNA-seq samples. Nat Commun *14*,
793 7286. 10.1038/s41467-023-42841-y.
794 66. Koop, A.H., Kahrilas, P.J., Schauer, J., Pandolfino, J.E., and Carlson, D.A. (2023). The impact of
795 primary peristalsis, contractile reserve, and secondary peristalsis on esophageal clearance measured
796 by timed barium esophagogram. Neurogastroenterol Motil *35*, e14638. 10.1111/nmo.14638.
797 67. Wang, Q., Li, Y., Wu, M., Huang, S., Zhang, A., Zhang, Y., and Jia, Z. (2021). Targeting
798 microsomal prostaglandin E synthase 1 to develop drugs treating the inflammatory diseases. Am J
799 Transl Res *13*, 391-419.
800 68. Yi, Y.S. (2016). Functional Role of Milk Fat Globule-Epidermal Growth Factor VIII in
801 Macrophage-Mediated Inflammatory Responses and Inflammatory/Autoimmune Diseases.
802 Mediators Inflamm *2016*, 5628486. 10.1155/2016/5628486.
803 69. Oh, S., Bournique, E., Bowen, D., Jalili, P., Sanchez, A., Ward, I., Dananberg, A., Manjunath, L.,
804 Tran, G.P., Semler, B.L., et al. (2021). Genotoxic stress and viral infection induce transient
805 expression of APOBEC3A and pro-inflammatory genes through two distinct pathways. Nat
806 Commun *12*, 4917. 10.1038/s41467-021-25203-4.
807 70. Tiwari, R., de la Torre, J.C., McGavern, D.B., and Nayak, D. (2019). Beyond Tethering the Viral
808 Particles: Immunomodulatory Functions of Tetherin (BST-2). DNA Cell Biol *38*, 1170-1177.
809 10.1089/dna.2019.4777.
810 71. Baaten, B.J., Li, C.R., and Bradley, L.M. (2010). Multifaceted regulation of T cells by CD44.
811 Commun Integr Biol *3*, 508-512. 10.4161/cib.3.6.13495.
812 72. Liu, Q., Niu, X., Li, Y., Zhang, J.R., Zhu, S.J., Yang, Q.Y., Zhang, W., and Gong, L. (2022). Role
813 of the mucin-like glycoprotein FCGBP in mucosal immunity and cancer. Front Immunol *13*,
814 863317. 10.3389/fimmu.2022.863317.
815 73. Okamoto, Y., and Shikano, S. (2023). Emerging roles of a chemoattractant receptor GPR15 and
816 ligands in pathophysiology. Front Immunol *14*, 1179456. 10.3389/fimmu.2023.1179456.
817 74. Subramanian, B.C., Moissoglu, K., and Parent, C.A. (2018). The LTB(4)-BLT1 axis regulates the
818 polarized trafficking of chemoattractant GPCRs during neutrophil chemotaxis. J Cell Sci *131*.
819 10.1242/jcs.217422.
820 75. Nguyen, H.T., Reyes-Alcaraz, A., Yong, H.J., Nguyen, L.P., Park, H.K., Inoue, A., Lee, C.S.,
821 Seong, J.Y., and Hwang, J.I. (2020). CXCR7: a beta-arrestin-biased receptor that potentiates cell
822 migration and recruits beta-arrestin2 exclusively through Gbetagamma subunits and GRK2. Cell
823 Biosci *10*, 134. 10.1186/s13578-020-00497-x.
824 76. Kelly, A., and Trowsdale, J. (2019). Genetics of antigen processing and presentation.
825 Immunogenetics *71*, 161-170. 10.1007/s00251-018-1082-2.
826 77. Bartel, Y., Bauer, B., and Steinle, A. (2013). Modulation of NK cell function by genetically
827 coupled C-type lectin-like receptor/ligand pairs encoded in the human natural killer gene complex.
828 Front Immunol *4*, 362. 10.3389/fimmu.2013.00362.
829 78. Lu, R.Q., Zhang, Y.Y., Zhao, H.Q., Guo, R.Q., Jiang, Z.X., and Guo, R. (2022). SGK1, a Critical
830 Regulator of Immune Modulation and Fibrosis and a Potential Therapeutic Target in Chronic Graft-
831 Versus-Host Disease. Front Immunol *13*, 822303. 10.3389/fimmu.2022.822303.
832 79. Li, Y., Su, G., Zhong, Y., Xiong, Z., Huang, T., Quan, J., Huang, J., Wen, X., Luo, C., Zheng, W.,
833 et al. (2021). HB-EGF-induced IL-8 secretion from airway epithelium leads to lung fibroblast
834 proliferation and migration. BMC Pulm Med *21*, 347. 10.1186/s12890-021-01726-w.
835 80. Uhlen, M., Fagerberg, L., Hallstrom, B.M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson,
836 A., Kampf, C., Sjostedt, E., Asplund, A., et al. (2015). Proteomics. Tissue-based map of the human
837 proteome. Science *347*, 1260419. 10.1126/science.1260419.

838 81. Politi, E., Angelakopoulou, A., Grapsa, D., Zande, M., Stefanaki, K., Panagiotou, I., Roma, E., and
839 Syrigou, E. (2017). Filaggrin and Periostin Expression Is Altered in Eosinophilic Esophagitis and
840 Normalized With Treatment. J Pediatr Gastroenterol Nutr *65*, 47-52.
841 10.1097/MPG.0000000000001419.
842 82. Oshima, N., Ishihara, S., Fukuba, N., Mishima, Y., Kawashima, K., Ishimura, N., Ishikawa, N.,
843 Maruyama, R., and Kinoshita, Y. (2017). Epidermal differentiation complex protein involucrin is
844 down-regulated in eosinophilic esophagitis. Esophagus *14*, 171-177. 10.1007/s10388-016-0568-y.
845 83. Yu, T., Yang, X., Fu, Q., Liang, J., Wu, X., Sheng, J., Chen, Y., Xiao, L., Wu, Y., Nie, D., et al.
846 (2023). TRIM11 attenuates Treg cell differentiation by p62-selective autophagic degradation of
847 AIM2. Cell Rep *42*, 113231. 10.1016/j.celrep.2023.113231.
848 84. Crowell, M.D., Umar, S.B., Griffing, W.L., DiBaise, J.K., Lacy, B.E., and Vela, M.F. (2017).
849 Esophageal Motor Abnormalities in Patients With Scleroderma: Heterogeneity, Risk Factors, and
850 Effects on Quality of Life. Clin Gastroenterol Hepatol *15*, 207-213 e201.
851 10.1016/j.cgh.2016.08.034.
852 85. Forbes, A., and Marie, I. (2009). Gastrointestinal complications: the most frequent internal
853 complications of systemic sclerosis. Rheumatology (Oxford) *48 Suppl 3*, iii36-39.
854 10.1093/rheumatology/ken485.
855 86. Krieg, T., Abraham, D., and Lafyatis, R. (2007). Fibrosis in connective tissue disease: the role of
856 the myofibroblast and fibroblast-epithelial cell interactions. Arthritis Res Ther *9 Suppl 2*, S4.
857 10.1186/ar2188.
858 87. Aden, N., Shiwen, X., Aden, D., Black, C., Nuttall, A., Denton, C.P., Leask, A., Abraham, D., and
859 Stratton, R. (2008). Proteomic analysis of scleroderma lesional skin reveals activated wound
860 healing phenotype of epidermal cell layer. Rheumatology (Oxford) *47*, 1754-1760.
861 10.1093/rheumatology/ken370.
862 88. Aden, N., Nuttall, A., Shiwen, X., de Winter, P., Leask, A., Black, C.M., Denton, C.P., Abraham,
863 D.J., and Stratton, R.J. (2010). Epithelial cells promote fibroblast activation via IL-1alpha in
864 systemic sclerosis. J Invest Dermatol *130*, 2191-2200. 10.1038/jid.2010.120.
865 89. Nikitorowicz-Buniak, J., Denton, C.P., Abraham, D., and Stratton, R. (2015). Partially Evoked
866 Epithelial-Mesenchymal Transition (EMT) Is Associated with Increased TGFbeta Signaling within
867 Lesional Scleroderma Skin. PLoS One *10*, e0134092. 10.1371/journal.pone.0134092.
868 90. Miyagawa, T., Asano, Y., Saigusa, R., Hirabayashi, M., Yamashita, T., Taniguchi, T., Takahashi,
869 T., Nakamura, K., Miura, S., Yoshizaki, A., et al. (2019). A potential contribution of trappin-2 to
870 the development of vasculopathy in systemic sclerosis. J Eur Acad Dermatol Venereol *33*, 753-760.
871 10.1111/jdv.15387.
872 91. Deraison, C., Bonnart, C., Langella, P., Roget, K., and Vergnolle, N. (2023). Elafin and its
873 precursor trappin-2: What is their therapeutic potential for intestinal diseases? Br J Pharmacol *180*,
874 144-160. 10.1111/bph.15985.
875 92. Hinchcliff, M., Huang, C.C., Wood, T.A., Matthew Mahoney, J., Martyanov, V., Bhattacharyya, S.,
876 Tamaki, Z., Lee, J., Carns, M., Podlusky, S., et al. (2013). Molecular signatures in skin associated
877 with clinical improvement during mycophenolate treatment in systemic sclerosis. J Invest Dermatol
878 *133*, 1979-1989. 10.1038/jid.2013.130.
879 93. Dong, X., Liu, Z., Lan, D., Niu, J., Miao, J., Yang, G., Zhang, F., Sun, Y., Wang, K., and Miao, Y.
880 (2017). Critical role of Keratin 1 in maintaining epithelial barrier and correlation of its down-
881 regulation with the progression of inflammatory bowel disease. Gene *608*, 13-19.
882 10.1016/j.gene.2017.01.015.
883 94. McCann, M.R., Monemdjou, R., Ghassemi-Kakroodi, P., Fahmi, H., Perez, G., Liu, S., Shi-Wen,
884 X., Parapuram, S.K., Kojima, F., Denton, C.P., et al. (2011). mPGES-1 null mice are resistant to
885 bleomycin-induced skin fibrosis. Arthritis Res Ther *13*, R6. 10.1186/ar3226.
886 95. Villanueva-Martin, G., Acosta-Herrera, M., Carmona, E.G., Kerick, M., Ortego-Centeno, N.,
887 Callejas-Rubio, J.L., Mages, N., Klages, S., Borno, S., Timmermann, B., et al. (2023). Non-
888 classical circulating monocytes expressing high levels of microsomal prostaglandin E2 synthase-1

889 tag an aberrant IFN-response in systemic sclerosis. J Autoimmun *140*, 103097.
890 10.1016/j.jaut.2023.103097.
891 96. Unlu, B., Tursen, U., Rajabi, Z., Jabalameli, N., and Rajabi, F. (2022). The Immunogenetics of
892 Systemic Sclerosis. Adv Exp Med Biol *1367*, 259-298. 10.1007/978-3-030-92616-8_10.
893 97. Liang, M., Lv, J., Jiang, Z., He, H., Chen, C., Xiong, Y., Zhu, X., Xue, Y., Yu, Y., Yang, S., et al.
894 (2020). Promotion of Myofibroblast Differentiation and Tissue Fibrosis by the Leukotriene B(4) -
895 Leukotriene B(4) Receptor Axis in Systemic Sclerosis. Arthritis Rheumatol *72*, 1013-1025.
896 10.1002/art.41192.
897 98. Hirabayashi, M., Asano, Y., Yamashita, T., Miura, S., Nakamura, K., Taniguchi, T., Saigusa, R.,
898 Takahashi, T., Ichimura, Y., Miyagawa, T., et al. (2018). Possible pro-inflammatory role of
899 heparin-binding epidermal growth factor-like growth factor in the active phase of systemic
900 sclerosis. J Dermatol *45*, 182-188. 10.1111/1346-8138.14088.
901 99. Guo, S., Li, Y., Wang, Y., Chu, H., Chen, Y., Liu, Q., Guo, G., Tu, W., Wu, W., Zou, H., et al.
902 (2016). Copy Number Variation of HLA-DQA1 and APOBEC3A/3B Contribute to the
903 Susceptibility of Systemic Sclerosis in the Chinese Han Population. J Rheumatol *43*, 880-886.
904 10.3899/jrheum.150945.
905 100. Lopez-Isac, E., Acosta-Herrera, M., Kerick, M., Assassi, S., Satpathy, A.T., Granja, J., Mumbach,
906 M.R., Beretta, L., Simeon, C.P., Carreira, P., et al. (2019). GWAS for systemic sclerosis identifies
907 multiple risk loci and highlights fibrotic and vasculopathy pathways. Nat Commun *10*, 4955.
908 10.1038/s41467-019-12760-y.
909 101. Acosta-Herrera, M., Kerick, M., Lopez-Isac, E., Assassi, S., Beretta, L., Simeon-Aznar, C.P.,
910 Ortego-Centeno, N., International, S.G., Proudman, S.M., Australian Scleroderma Interest, G., et al.
911 (2021). Comprehensive analysis of the major histocompatibility complex in systemic sclerosis
912 identifies differential HLA associations by clinical and serological subtypes. Ann Rheum Dis *80*,
913 1040-1047. 10.1136/annrheumdis-2021-219884.
914 102. Hanson, A.L., Sahhar, J., Ngian, G.S., Roddy, J., Walker, J., Stevens, W., Nikpour, M., Assassi, S.,
915 Proudman, S., Mayes, M.D., et al. (2022). Contribution of HLA and KIR Alleles to Systemic
916 Sclerosis Susceptibility and Immunological and Clinical Disease Subtypes. Front Genet *13*,
917 913196. 10.3389/fgene.2022.913196.
918 103. Dai, H., Wang, L., Li, L., Huang, Z., and Ye, L. (2021). Metallothionein 1: A New Spotlight on
919 Inflammatory Diseases. Front Immunol *12*, 739918. 10.3389/fimmu.2021.739918.
920 104. Subramanian Vignesh, K., and Deepe, G.S., Jr. (2017). Metallothioneins: Emerging Modulators in
921 Immunity and Infection. Int J Mol Sci *18*. 10.3390/ijms18102197.
922 105. Marie, I. (2019). Systemic sclerosis and exposure to heavy metals. Autoimmun Rev *18*, 62-72.
923 10.1016/j.autrev.2018.11.001.
924 106. Shaker, R., Dodds, W.J., Helm, J.F., Kern, M.K., and Hogan, W.J. (1991). Regional esophageal
925 distribution and clearance of refluxed gastric acid. Gastroenterology *101*, 355-359. 10.1016/0016-
926 5085(91)90011-9.
927 107. Diener, U., Patti, M.G., Molena, D., Fisichella, P.M., and Way, L.W. (2001). Esophageal
928 dysmotility and gastroesophageal reflux disease. J Gastrointest Surg *5*, 260-265. 10.1016/s1091-
929 255x(01)80046-9.
930 108. Patel, C.K., Kahrilas, P.J., Hodge, N.B., Tsikretsis, L.E., Carlson, D.A., Pandolfino, J.E., and
931 Tetreault, M.P. (2022). RNA-sequencing reveals molecular and regional differences in the
932 esophageal mucosa of achalasia patients. Sci Rep *12*, 20616. 10.1038/s41598-022-25103-7.
933 109. Woodland, P., Aktar, R., Mthunzi, E., Lee, C., Peiris, M., Preston, S.L., Blackshaw, L.A., and
934 Sifrim, D. (2015). Distinct afferent innervation patterns within the human proximal and distal
935 esophageal mucosa. Am J Physiol Gastrointest Liver Physiol *308*, G525-531.
936 10.1152/ajpgi.00175.2014.
937 110. Sandqvist, G., and Hesselstrand, R. (2019). Validity of the Swedish version of the systemic
938 sclerosis quality of life questionnaire (SSCQoL): A novel measure of quality of life for patients
939 with systemic sclerosis. Ann Rheum Dis *78*, 855-857. 10.1136/annrheumdis-2018-214260.

940 111. Lee, A.S., Lee, J.S., He, Z., and Ryu, J.H. (2020). Reflux-Aspiration in Chronic Lung Disease. Ann
941 Am Thorac Soc *17*, 155-164. 10.1513/AnnalsATS.201906-427CME.
942 112. Savarino, E., Bazzica, M., Zentilin, P., Pohl, D., Parodi, A., Cittadini, G., Negrini, S., Indiveri, F.,
943 Tutuian, R., Savarino, V., and Ghio, M. (2009). Gastroesophageal reflux and pulmonary fibrosis in
944 scleroderma: a study using pH-impedance monitoring. Am J Respir Crit Care Med *179*, 408-413.
945 10.1164/rccm.200808-1359OC.
946 113. Conboy, C.M., Spyrou, C., Thorne, N.P., Wade, E.J., Barbosa-Morais, N.L., Wilson, M.D.,
947 Bhattacharjee, A., Young, R.A., Tavare, S., Lees, J.A., and Odom, D.T. (2007). Cell cycle genes
948 are the evolutionarily conserved targets of the E2F4 transcription factor. PLoS One *2*, e1061.
949 10.1371/journal.pone.0001061.
950 114. Garcia-Gutierrez, L., Delgado, M.D., and Leon, J. (2019). MYC Oncogene Contributions to
951 Release of Cell Cycle Brakes. Genes (Basel) *10*. 10.3390/genes10030244.
952 115. Lastra, D., Escoll, M., and Cuadrado, A. (2022). Transcription Factor NRF2 Participates in Cell
953 Cycle Progression at the Level of G1/S and Mitotic Checkpoints. Antioxidants (Basel) *11*.
954 10.3390/antiox11050946.
955 116. Portal, C., Wang, Z., Scott, D.K., Wolosin, J.M., and Iomini, C. (2022). The c-Myc Oncogene
956 Maintains Corneal Epithelial Architecture at Homeostasis, Modulates p63 Expression, and
957 Enhances Proliferation During Tissue Repair. Invest Ophthalmol Vis Sci *63*, 3. 10.1167/iovs.63.2.3.
958 117. Wruck, C.J., Wruck, A., Brandenburg, L.O., Kadyrov, M., Tohidnezhad, M., and Pufe, T. (2011).
959 Impact of Nrf2 on esophagus epithelium cornification. Int J Dermatol *50*, 1362-1365.
960 10.1111/j.1365-4632.2011.04989.x.
961 118. Jones, R., Junghard, O., Dent, J., Vakil, N., Halling, K., Wernersson, B., and Lind, T. (2009).
962 Development of the GerdQ, a tool for the diagnosis and management of gastro-oesophageal reflux
963 disease in primary care. Aliment Pharmacol Ther *30*, 1030-1038. 10.1111/j.1365-
964 2036.2009.04142.x.
965

966 **FIGURES**

967 **Figure 1. ScRNA-seq sample composition by cell type. A)** Integrated UMAP embedding of all cells
968 (n=306,372) labeled by cell type. **B)** Expression by cell type of canonical markers. **C)** Sample-wise (n=39)
969 distribution of major cell types. **D)** Proportion of cells by cell type.

970

971 **Figure 2. Landscape of esophageal epithelial cells. A)** Histological summary of the human esophageal
972 epithelium, adopted from Clevenger et al.[28] Alternatively shaded cells within the same layer represent
973 proliferating cells. NOTE: The layers of replicating cells not attached to the basement membrane, labeled
974 here as proliferating suprabasal cells, are sometimes referred to as epibasal cells[28]. **B)** Integrated UMAP
975 embedding of all esophageal epithelial cells (EECs, n=230,720) labeled by epithelial compartment. **C)**
976 Expression by EEC compartment and cell cycle markers. **C)** Proportion of proliferating basal and
977 suprabasal cells by condition and biopsy location. **E)** Proportions of cells by condition, EEC compartment,
978 and biopsy location. The bars denote the mean values, the vertical lines the standard deviations, and the
979 points the individual sample proportions. Pairwise differences across conditions were evaluated statistically
980 using MASC[52]. *$p<0.05$; **$p<0.005$.

981

982 **Figure 3. Distribution of esophageal epithelial cells by differentiation score and epithelial**
983 **compartment. A)** Integrated UMAP embeddings of esophageal epithelial cells divided by disease state,
984 colored by differentiation score. **B)** Swarm plots of esophageal epithelial cells by differentiation score,
985 colored by epithelial compartment.

986

987 **Figure 4. Differential gene expression at single-cell resolution in esophageal epithelial cells. A)** Inter-
988 sample expression correlations are shown for each epithelial compartment for the 2,000 most variable
989 genes. **B)** The distribution of inter-sample expression correlations are shown by epithelial compartment. **C-**
990 **D)** UpSet plots showing differential gene expression at the single-cell level by epithelial compartment for
991 SSc vs HCs (purple) and GERD vs HCs (orange) in the proximal **(C)** and distal **(D)** esophagus. **E)** The
992 number of significantly differentially expressed genes with $|\log_2 FC| > 0.1$ are shown for SSC vs. HCs and
993 GERD vs. HCs, by non-proliferating epithelial compartment and biopsy location. **F)** Gene set enrichment

994 analysis results, showing all pathways enriched with P<0.001 (unadjusted) for SSc vs. HCs (purple) and
995 GERD vs. HCs (orange), split by non-proliferating epithelial compartment and biopsy location. Pathways in
996 bold had enrichment with P<0.001 in both SSc and GERD. Points circled in red indicate statistical
997 significance after adjusting for the number of tested pathways (P_{FDR}<0.05). **G-H)** Relative gene expression
998 (SSc vs. HCs) for all DEGs at the single-cell level were modeled using Lamian[65] and plotted against the
999 epithelial differentiation score in the proximal **(G)** and distal **(H)** esophagus. These trajectories were
1000 clustered based on Euclidean distance using Ward's method. The lefthand color bar displays the blended
1001 color of all epithelial compartments in which each gene was differentially expressed. In **G)**, labeled genes
1002 highlight the top 25 ranked genes in the leading edge of the innate immune system pathway enriched in
1003 SSc in the superficial compartment of the proximal esophagus. In **H)**, the labeled genes mark the leading
1004 edge genes of the matrisome pathway enriched in SSc in the suprabasal compartment of the distal
1005 esophagus. **I-L)** Dot plots showing average gene expression, by condition and epithelial compartment, for
1006 the most highly represented protein families in the leading edges of significantly enriched pathways:
1007 keratins **(I)**, S100 proteins **(J)**, small proline-rich proteins **(K)**, and serine protease inhibitors **(L)**. **M)** Single-
1008 cell gene expression by condition, epithelial compartment, and esophageal region for *PI3*, the most frequent
1009 gene in leading edges of significantly enriched pathways.
1010
1011 **Figure 5. Differential gene expression aggregated by sample and EEC compartment. A-B)** Gene
1012 expression differences between SSc and HCs **(A)** and between SSc and GERD **(B)** are shown for each
1013 EEC compartment. Significant DEGs (FDR<0.05) are highlighted in purple. The percentages of genes that
1014 were significantly differentially expressed in each compartment are plotted in blue in along the secondary
1015 Y axis. **C-D)** Gene expression differences in log_2FC are shown for SSc vs. HCs (Y axis) against GERD vs.
1016 HCs (X axis) within the superficial compartment in the proximal **(C)** and distal **(D)** regions. Significant DEGs
1017 in both conditions are highlighted in blue, significant DEGs in SSC only are highlighted in purple, and
1018 significant DEGs in GERD only are highlighted in orange. The correlation between comparisons is shown
1019 in the upper left, and the slope and coefficient of determination for the modeled linear regression with
1020 intercept=0 is displayed in the lower right. Trendlines with 95% confidence intervals are plotted in pink, and
1021 the dashed grey lines denote where y=x. Encircled points highlight disease-specific DEGs that are

1022 referenced in the main text and plotted in **(E)**, which shows the distribution of expression aggregated by
1023 sample in the superficial compartment across conditions and biopsy locations. Pairwise differential
1024 expression was determined using edgeR's quasi-likelihood F-tests: *p<0.05; **p<0.05; ***p<0.0005. **F)** The
1025 number of significant DEGs are shown for GERD vs. HCs (orange) and for all permutations of n=4 samples
1026 for SSc vs. HCs (purple) in superficial cells for both proximal and distal regions. The point within the SSc
1027 distribution is the median value. **G-H)** The distributions of relative FC or slope *m* **(G)** and coefficient of
1028 determination r^2 **(H)** are shown for all permutations of n=4 SSc samples, determined by modeling a linear
1029 regression with intercept=0 for the $\log_2$FC of SSc vs. HCs against the $\log_2$FC of GERD vs. HCs for all
1030 expressed genes in the superficial compartment for both proximal and distal regions. DEG, differentially
1031 expressed gene; FC, fold change.
1032
1033 **Figure 6. Landscape of superficial esophageal epithelial cells. A)** Integrated UMAP embedding of all
1034 superficial EECs (n=30,866) clustered by unique transcriptional signatures. **B)** Relative expression of
1035 cluster-specific gene signatures. **C)** Epithelial differentiation score by superficial cell cluster. **D)** Proportion
1036 of superficial cells by condition and cluster from the proximal esophagus. The bars denote the mean values,
1037 the vertical lines the standard deviations, and the points the individual sample proportions. Pairwise
1038 differences across conditions were evaluated statistically using MASC[52]. *p<0.05; **p<0.005. **E)** Distribution
1039 of metallothionein module score in proximal, superficial cells by condition. The metallothionein module
1040 score included *MT1A*, *MT1E*, *MT1F*, *MT1G*, *MT1H*, *MT1M*, *MT1X*, and *MT2A*. **F)** Distribution of relative
1041 metallothionein expression by cluster and condition in superficial cells from proximal esophagus. All
1042 pairwise comparisons were between SSc and GERD were statistically significant, except for cluster 5.
1043
1044 **Figure 7. Expression of enriched transcription factors and their targets in superficial EECs in**
1045 **proximal esophagus.** The distributions of enriched transcription factor expression and target module
1046 scores are shown proximal, superficial cells by condition and superficial cluster for **A)** IRF1, **B)** MYC, **C)**
1047 E2F4, and **D)** NE2L2.
1048

1049 **Figure 8. Correlations with clinical phenotypes. A)** Heatmap of quantitative clinical traits organized using
1050 agglomerative, complete hierarchical clustering on Euclidean distances, with pairwise Spearman
1051 correlations shown in each cell. **B)** SSc cases are plotted on the first two PCs of the quantitative clinical
1052 traits, colored by esophageal motility phenotype. The relative magnitude and direction of trait correlations
1053 with the PCs are shown with black arrows. **C)** Correlations between EEC compartments and clinical trait
1054 PCs in SSc cases. Linear regression trendlines with 95% confidence intervals are shown for each
1055 compartment against PC1 and PC2. **D)** Correlations between aggregate, relative metallothionein
1056 expression and clinical trait PCs in superficial cells in the proximal esophagus. Trendlines with 95%
1057 confidence intervals are shown with the unadjusted correlation P-value. The metallothionein module score
1058 included *MT1A*, *MT1E*, *MT1F*, *MT1G*, *MT1H*, *MT1M*, *MT1X*, and *MT2A*. **E)** Correlations between superficial
1059 clusters (SCs) and clinical trait PC1 in SSc cases. Linear regression trendlines with 95% confidence
1060 intervals are shown for each SC. The displayed P-values correspond to the highlighted correlations
1061 between SC3 and PC1. **F)** Boxplots showing the proportion of SC3 in superficial cells by esophageal motility
1062 phenotype in the proximal and distal esophagus, colored by disease state (HCs, grey; GERD, orange; SSc,
1063 purple). AET, acid exposure time; BEDQ, brief esophageal dysphagia questionnaire; BMI, body mass index;
1064 DCI, distal contractile integral; DI, distensibility index; EGJ, esophagogastric junction; EGJOO,
1065 esophagogastric junction outflow obstruction; FLIP, functional luminal imaging probe panometry; HCs,
1066 healthy controls; HRM, high-resolution manometry; IEM, ineffective esophageal motility; IQR, inter-quartile
1067 range; IRP, integrated relaxation pressure; NEQOL, Northwestern esophageal quality of life; NM,
1068 neuromyogenic model. NR, not reported.

1069 **SUPPLEMENTAL FIGURES**

1070 **Supplemental Figure 1. Cell clustering and cell type annotation. A)** Cell type module scores for 10
1071 detected cell types, colored by relative gene expression of the listed genes. **B)** Projected UMAP
1072 embeddings calculated on the top 40 PCs with n=200 neighbors. Cells are colored by cluster, which are
1073 numbered in descending order by total cell count. Clusters were determined using Seurat's modularity-
1074 based clustering on a shared nearest neighbor (SNN) graph with k=25 and resolution=0.25 on the top 40
1075 PCs. **C)** Isolation of cluster 13 for annotation of cell types with smaller proportions of cells and projected
1076 UMAP embeddings for cluster 13, colored by subcluster. **D)** Cell type module scores for 7 detected cell
1077 types in subcluster 13, colored by relative gene expression of the listed genes.

1078

1079 **Supplemental Figure 2. Epithelial cell quality control and annotation. A)** Projected UMAP embeddings
1080 for integrated, esophageal epithelial cells (EECs) calculated on the top 35 PCs with n=100 neighbors. Cells
1081 are colored by cluster, which are numbered in descending order by total cell count. Clusters were
1082 determined using Seurat's modularity-based clustering on a shared nearest neighbor (SNN) graph with
1083 k=40 and resolution=0.50 on the top 40 PCs**. B)** EEC UMAP projection, colored by the number of unique
1084 genes per cell. **C)** Distribution of number of UMIs detected per cell, split by cluster. EECs with <3500
1085 UMIs/cell were removed from further analysis. **D)** EEC UMAPs colored by gene expression (log(counts+1)))
1086 for genes included in compartment-specific expression scores. **E)** EEC UMAPs colored by compartment
1087 expression scores. **F)** Projected UMAP embeddings for post-quality-control EECs calculated on the top 35
1088 PCs with n=100 neighbors. Cells are colored by cluster, which are numbered in descending order by total
1089 cell count. Clusters were determined using Seurat's modularity-based clustering on a SNN graph with k=50
1090 and resolution=0.25 on the top 35 PCs. **G)** Result of K-means clustering to distinguish suprabasal from
1091 basal cells, based on expression of corresponding cluster markers (Supplemental Figure 4) in clusters 3,
1092 4, and 5, with K=2. **H)** Result of K-means clustering to distinguish superficial from suprabasal cells, based
1093 on expression of corresponding cluster markers (Supplemental Figure 4) in clusters 6, 7, and 8, with K=2.
1094 **I)** Predicted classification of cell cycle state in EECs determined using using Seurat's CellCycleScoring()
1095 function.

1096

1097 **Supplemental Figure 3. Derivation of epithelial cell differentiation score. A)** Density plots of
1098 esophageal epithelial cell compartment module scores by compartment annotation. **B)** Density plot of
1099 epithelial cell differentiation score by annotated compartment. **C)** UMAP of esophageal epithelial cells
1100 colored by differentiation score.
1101
1102 **Supplemental Figure 4. Group-wise gene expression correlation for variable genes. A)** 2,000 most
1103 variable genes among single cells across all samples, with top 20 labeled. **B)** Correlations of gene
1104 expression aggregated by condition and biopsy location are shown for each epithelial compartment for the
1105 2,000 most variable genes, with proximal samples along the Y axis and distal samples along the X axis.
1106
1107 **Supplemental Figure 5. Differential gene expression aggregated by sample and esophageal**
1108 **epithelial cell compartment. A)** Multidimensional scaling plots of aggregate gene expression samples
1109 colored by sample, epithelial compartment, condition, and biopsy location. **B)** Volcano plots of differential
1110 gene expression between SSc and GERD by epithelial compartment and biopsy location with significant
1111 associations highlighted in red and outer significantly differentially expressed genes labeled. **C)** Gene
1112 expression differences in $\log_2$FC are shown for SSc vs. HCs (Y axis) against GERD vs. HCs (X axis) for
1113 each epithelial compartment in the proximal and distal regions. Outer significantly differentially expressed
1114 genes are labeled, as well as immune-related genes uniquely differentially expressed in SSc. The
1115 Spearman correlations between gene expression differences are shown in the upper right of each plot (*r*),
1116 the relative FC or slope (*m)* and coefficient of determination (r^2) are shown in the bottom right of each plot.
1117 Trendlines with 95% confidence intervals are plotted in pink, and the dashed grey lines denote where y=x.
1118 **D)** Distributions of aggregate gene expression by epithelial layer, biopsy location, and condition are shown
1119 for top SSc-specific and GERD-specific differentially expressed genes.
1120
1121 **Supplemental Figure 6. Expression of superficial esophageal epithelial cell markers.** UMAPs with
1122 expression for the top superficial cluster markers are shown along with the epithelial differentiation score
1123 for superficial cells.
1124

1125 **Supplemental Figure 7. *FLI1* expression in human esophageal mucosa. A)** UMAP of integrated
1126 mucosal dataset, colored by *FLI1* expression. The distribution of *FLI1* expression by cell type is shown in
1127 the upper right. **B)** Distribution of *FLI1* expression by cell type and disease state. Endothelial cells had the
1128 highest *FLI1* expression, but the expression differences between diseases was not significant. **C)** Violin
1129 plot of the relative expression of FLI1 target genes in superficial cells.
1130
1131 **Supplemental Figure 8. Epithelial cell compartment proportions and additional correlations by**
1132 **motility phenotype. A-B)** Proportions of esophageal epithelial cell compartments by HRM **(A)** and FLIP
1133 **(B)** esophageal dysmotility phenotypes are shown for the proximal and distal esophagus samples. The
1134 bars denote the mean values, the vertical lines the standard deviations, and the points the individual
1135 sample proportions. **C)** The relationship between the proportion of superficial cells and the SSc disease
1136 duration is plotted for proximal and distal samples. Trendlines with 95% confidence intervals are plotted in
1137 pink, and correlation P-values are displayed in the upper right corners. **D)** The distribution of normalized
1138 gene expression values by esophageal motility phenotype are shown for the *TRIM11* gene for the FLIP,
1139 HRM, and NM esophogeal dysmotility classifications. Unadjusted ordinal regression p-values are
1140 displayed in the upper left corners. *TRIM11* was the most strongly association gene with the dysmotility
1141 phenotypes by p-value. HRM, high-resolution manometry; FLIP, functional luminal imaging probe
1142 panometry; IEM, ineffective esophageal motility; EGJOO, esophagogastric junction outflow obstruction;
1143 NM, neuromyogenic model.

1144 **TABLES**

Table 1. 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[39] score	6	0 (0-0)	8	8 (3-14)	3	1 (1-1)
NEQOL[40] score	6	56 (56-56)	8	31 (23-47)	3	36 (29-46)

1145 AET, acid exposure time; BEDQ, brief esophageal dysphagia questionnaire; BMI, body mass index; DCI, distal contractile integral;
1146 DI, distensibility index; EGJ, esophagogastric junction; HCs, healthy controls; IQR, inter-quartile range; IRP, integrated relaxation
1147 pressure; NEQOL, Northwestern esophageal quality of life; NR, not reported

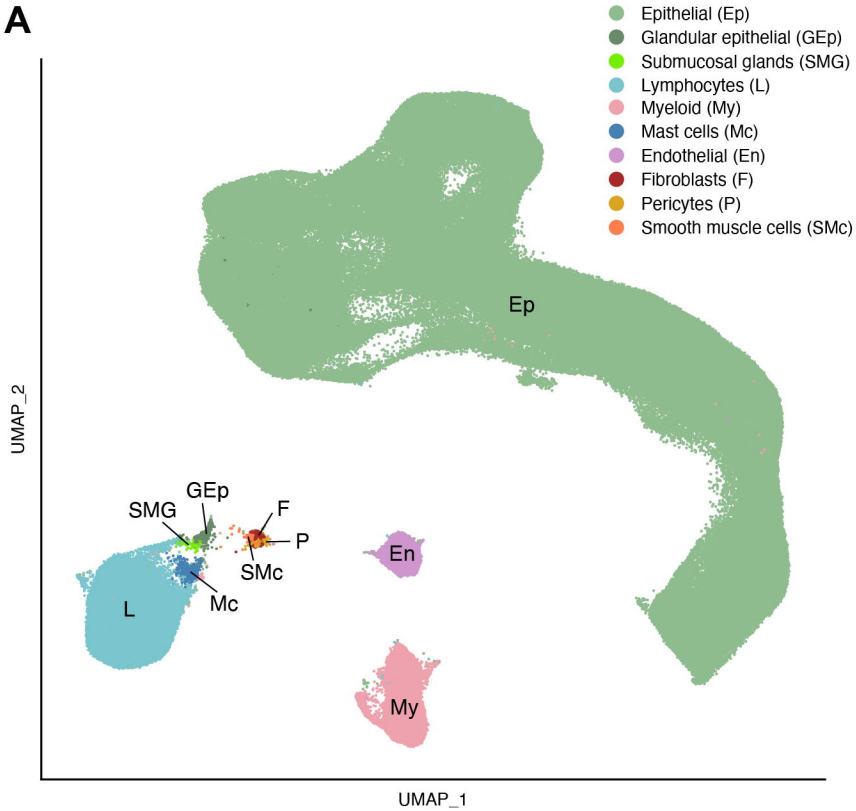
1148

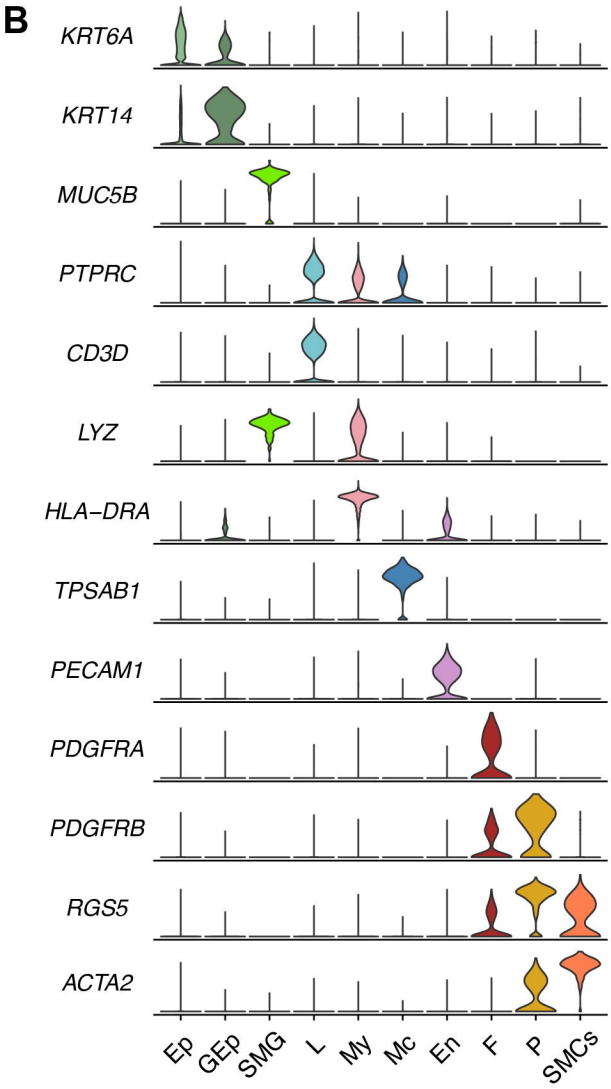
Table 2. Pairwise transcription factor enrichment in superficial cells in proximal esophagus

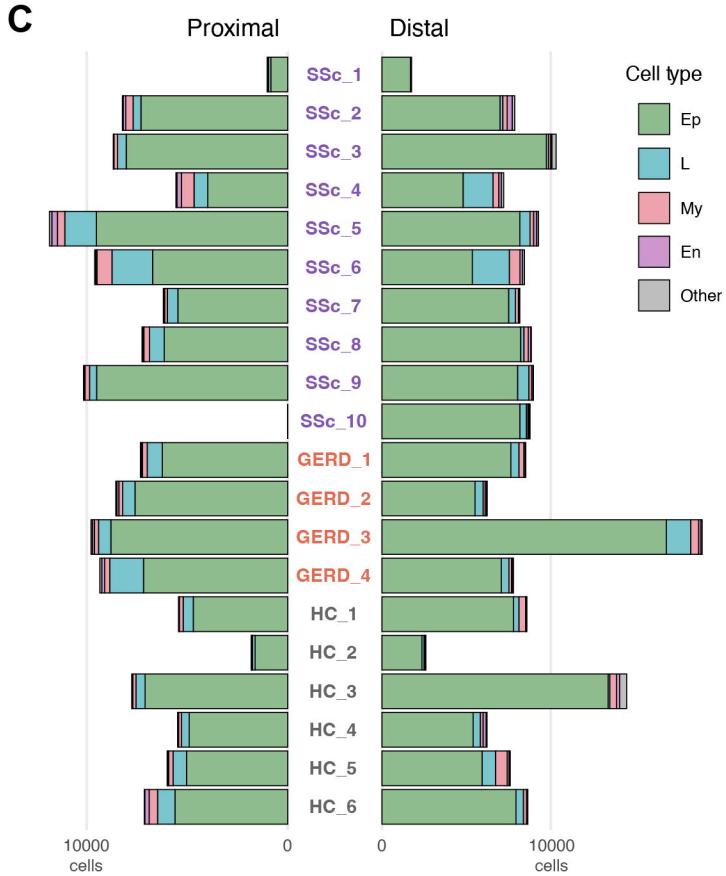
Comparison	TF	Direction	DEG Targets	DEG Targets / DEGs	TF Targets / Background	P_{adj}	TF Expr. LogFC	TF Expr. P_{adj}
SSc vs. **GERD**	IRF1	**+/-**	7	0.11	0.03	2.9×10^{-2}	0.22	9.0×10^{-1}
		-	5	0.19	0.03	8.9×10^{-3}		
SSc vs. **HCs**	MYC	**+**	120	0.10	0.06	2.1×10^{-7}	1.64	7.8×10^{-5}
	E2F4	**+**	116	0.08	0.06	5.5×10^{-5}	0.88	9.8×10^{-4}
	NFE2L2	**-**	108	0.11	0.07	7.6×10^{-4}	0.25	3.2×10^{-1}

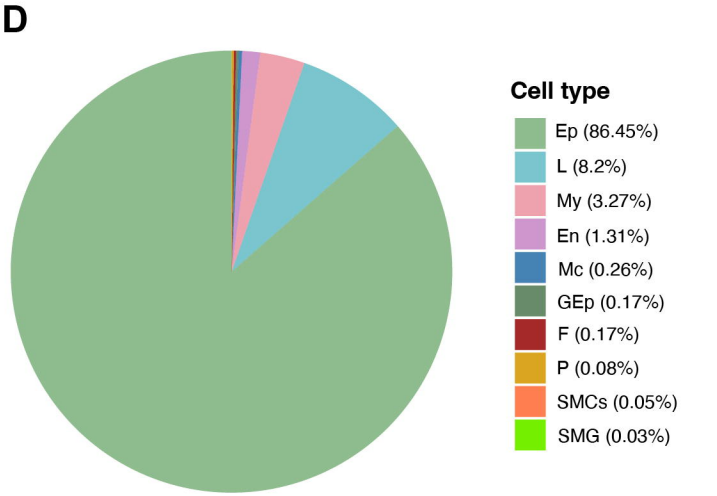
DEG, differentially expressed gene; Expr., expression; GERD, gastroesophageal reflux disease ; HCs, healthy controls; LogFC, $\log_2$ (fold change); SSc, systemic sclerosis; TF, transcription factor.

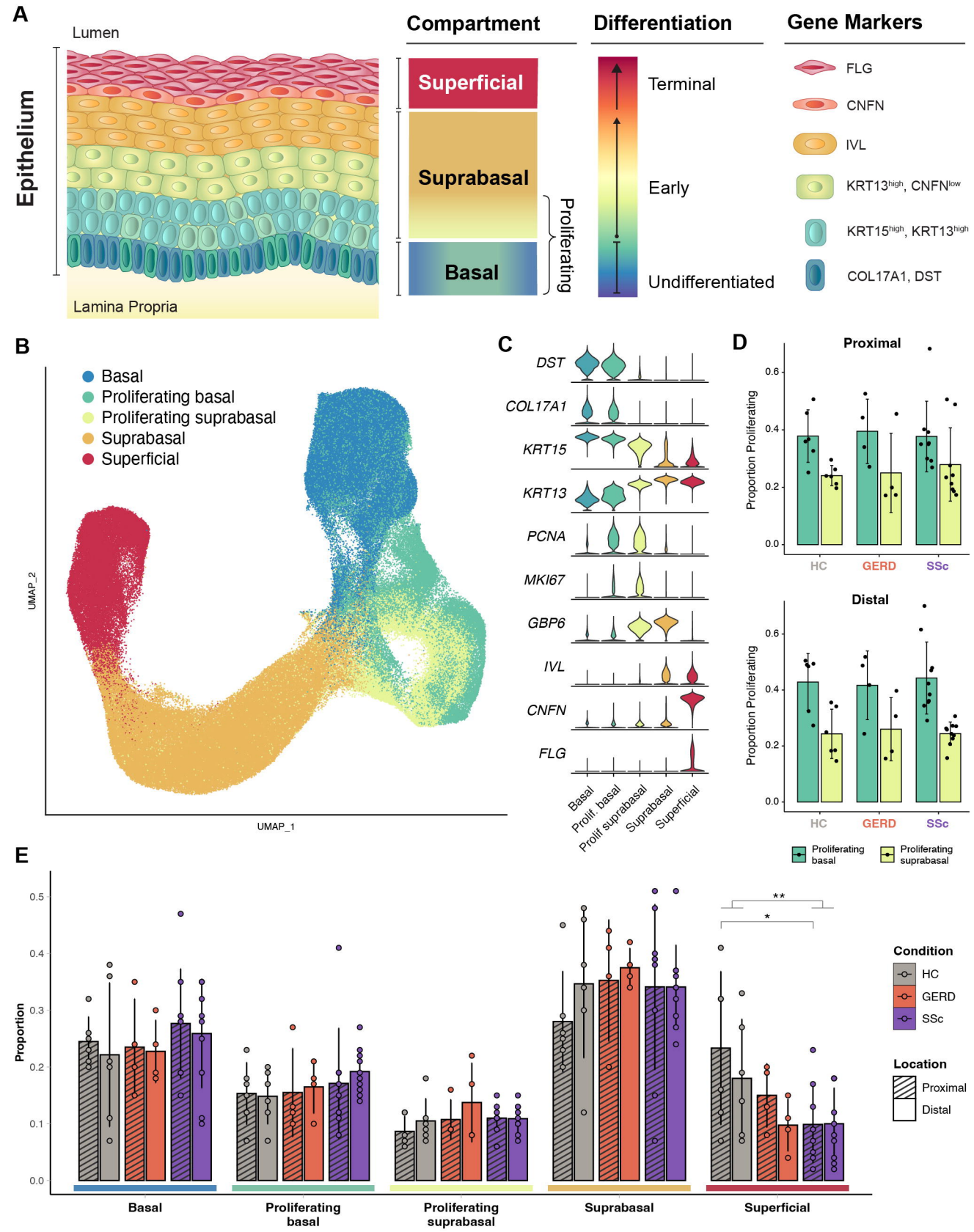
1149

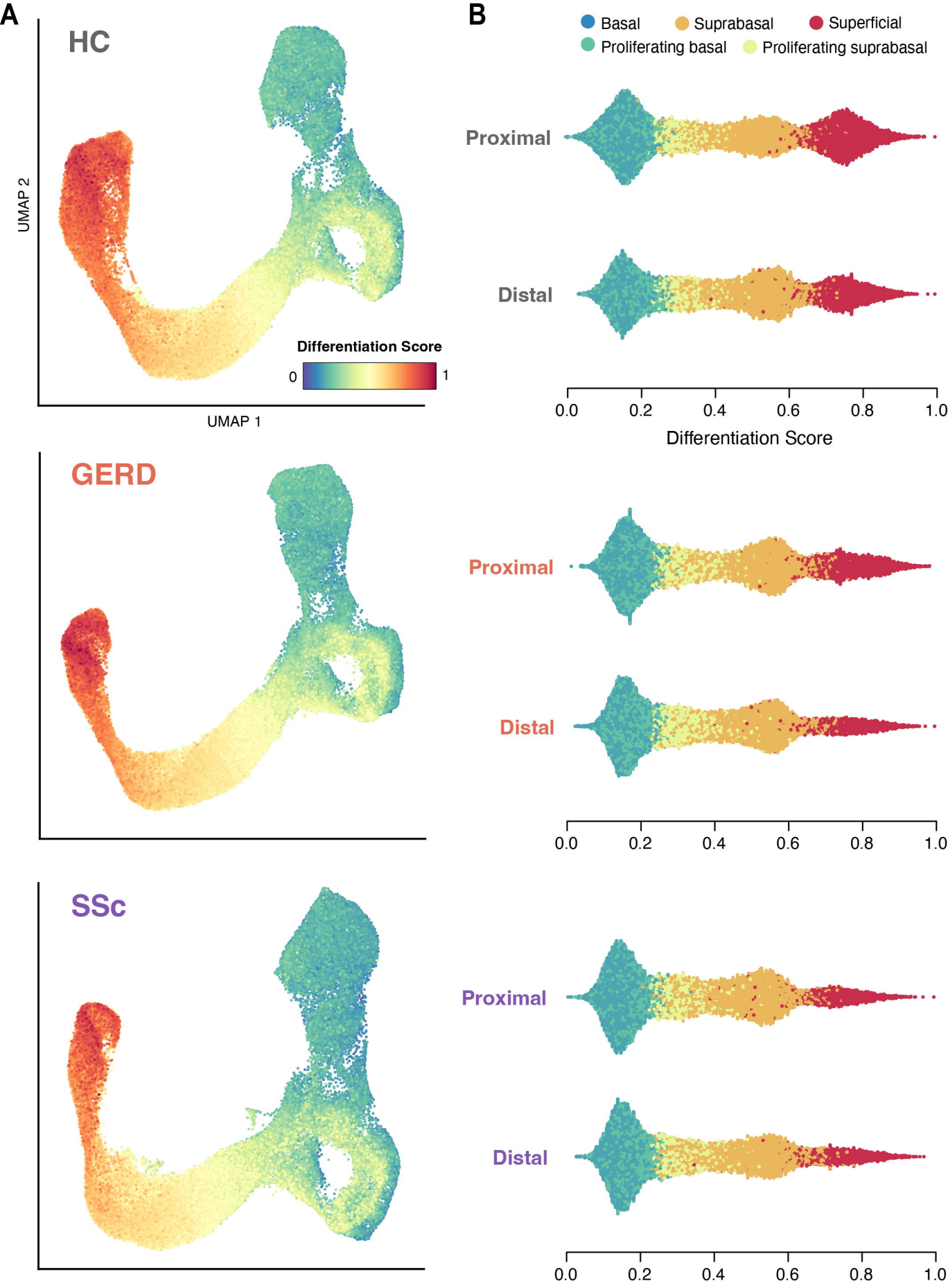

A
Epithelial (Ep)
Glandular epithelial (GEp)
Submucosal glands (SMG)
Lymphocytes (L)
Myeloid (My)
Mast cells (Mc)
Endothelial (En)
Fibroblasts (F)
Pericytes (P)
Smooth muscle cells (SMc)
Ep
GEp
SMG
F
P
SMc
Mc
L
En
My
UMAP_2
UMAP_1

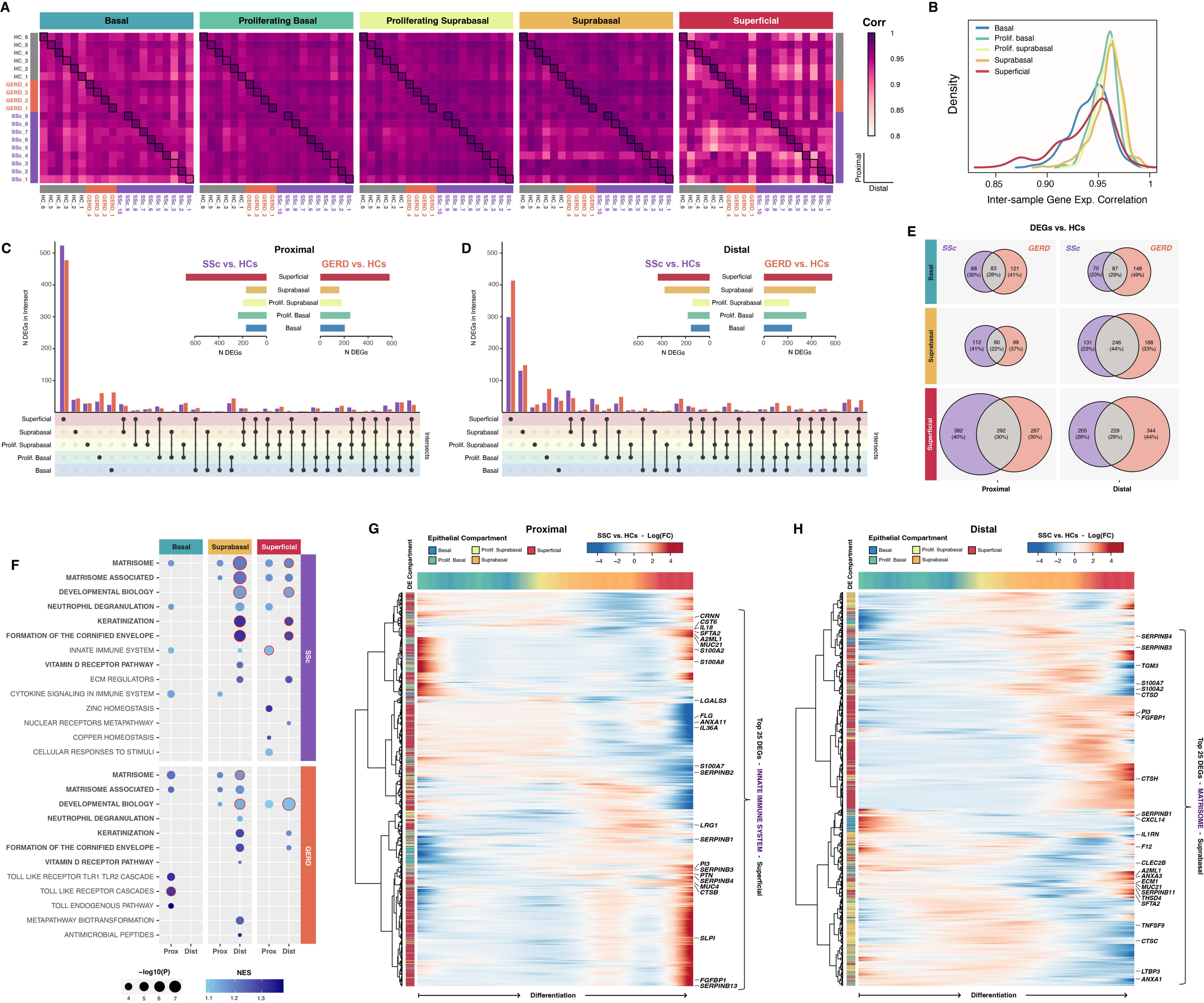

B
KRT6A
KRT14
MUC5B
PTPRC
CD3D
LYZ
HLA−DRA
TPSAB1
PECAM1
PDGFRA
PDGFRB
RGS5
ACTA2
Ep
GEp
SMG
L
My
Mc
En
F
P
SMCs

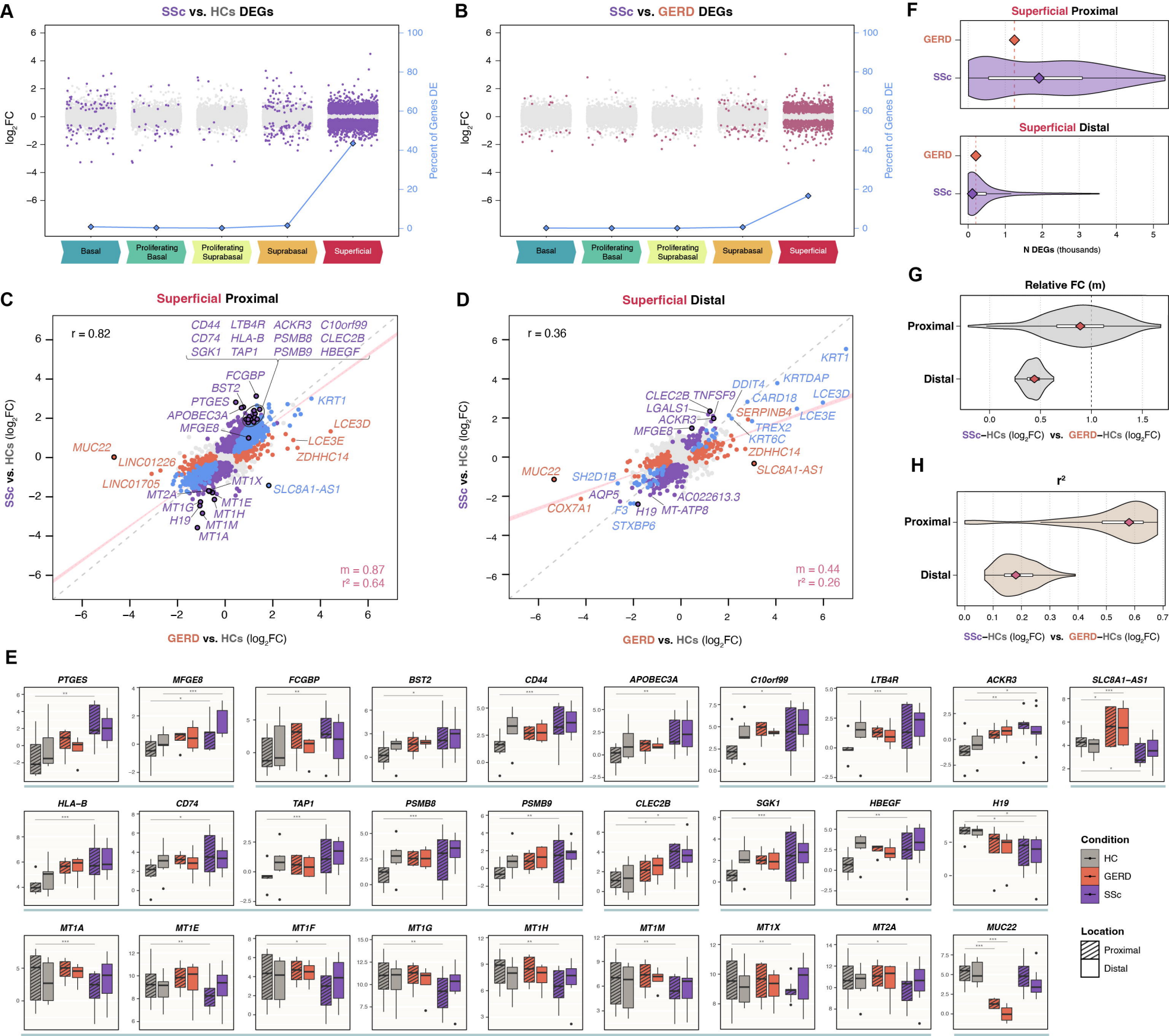

C
Proximal
Distal
SSc_1
SSc_2
SSc_3
SSc_4
SSc_5
SSc_6
SSc_7
SSc_8
SSc_9
SSc_10
GERD_1
GERD_2
GERD_3
GERD_4
HC_1
HC_2
HC_3
HC_4
HC_5
HC_6
Cell type
Ep
L
My
En
Other
10000
cells
0
0
10000
cells

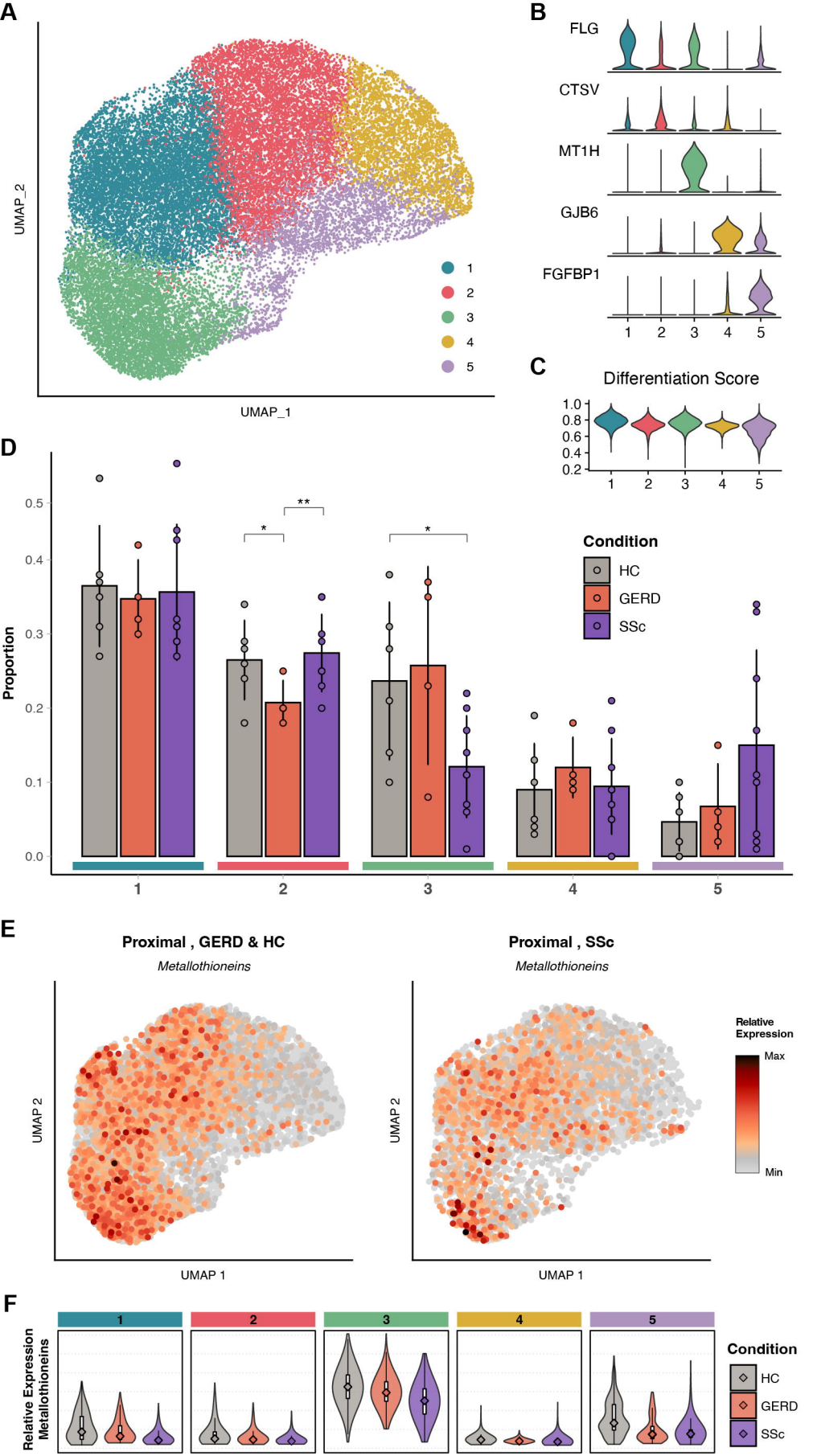

D
Cell type
Ep (86.45%)
L (8.2%)
My (3.27%)
En (1.31%)
Mc (0.26%)
GEp (0.17%)
F (0.17%)
P (0.08%)
SMCs (0.05%)
SMG (0.03%)

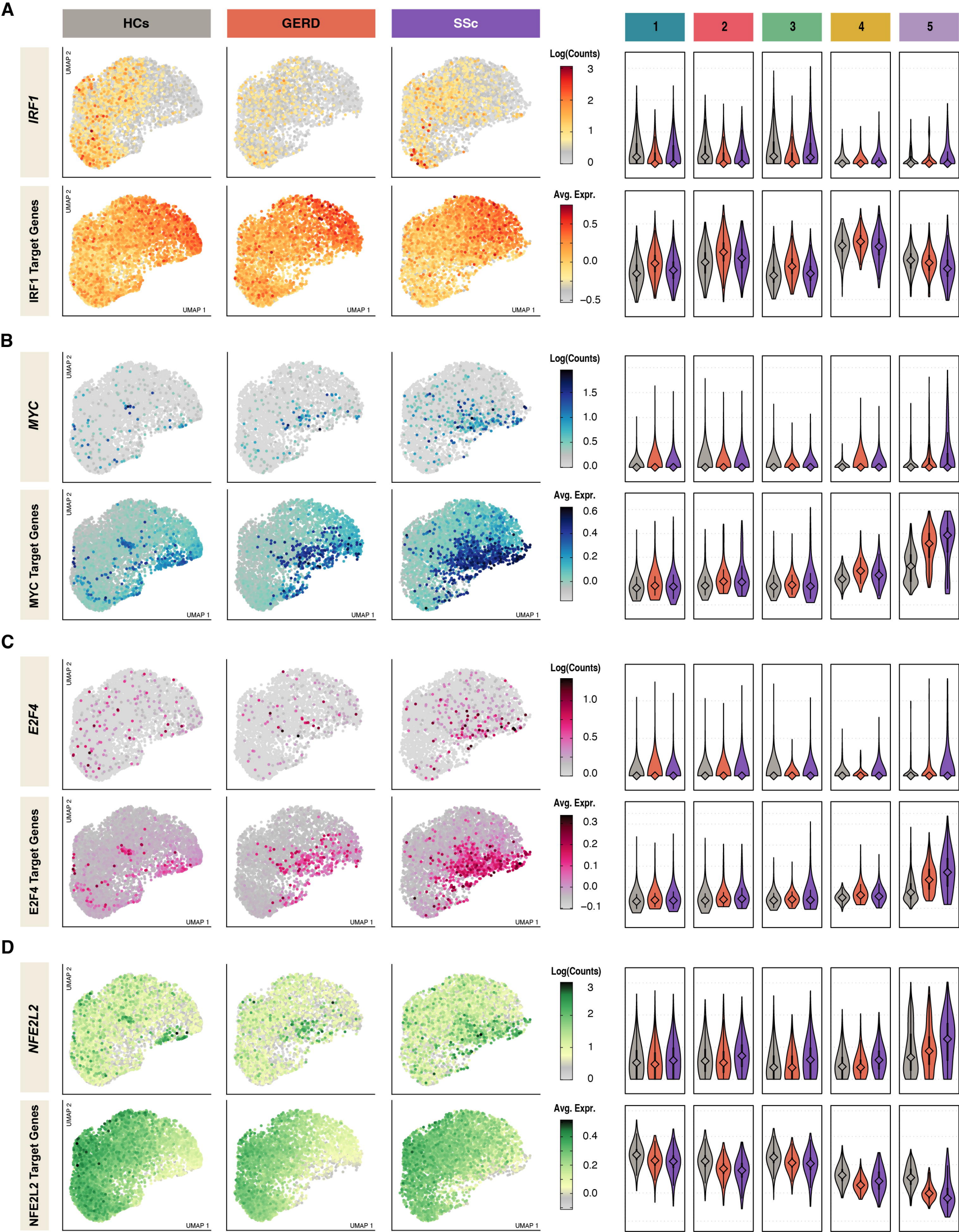

A
Lumen
Epithelium
Lamina Propria
Compartment
Superficial
Suprabasal
Basal
Proliferating
Differentiation
Terminal
Early
Undifferentiated
Gene Markers
FLG
CNFN
IVL
KRT13high, CNFNlow
KRT15high, KRT13high
COL17A1, DST
B
Basal
Proliferating basal
Proliferating suprabasal
Suprabasal
Superficial
UMAP_2
UMAP_1
C
DST
COL17A1
KRT15
KRT13
PCNA
MKI67
GBP6
IVL
CNFN
FLG
Basal
Prolif. basal
Prolif suprabasal
Suprabasal
Superficial
D
Proximal
Distal
Proportion Proliferating
HC
GERD
SSc
Proliferating basal
Proliferating suprabasal
E
Proportion
Basal
Proliferating basal
Proliferating suprabasal
Suprabasal
Superficial
**
*
Condition
HC
GERD
SSc
Location
Proximal
Distal

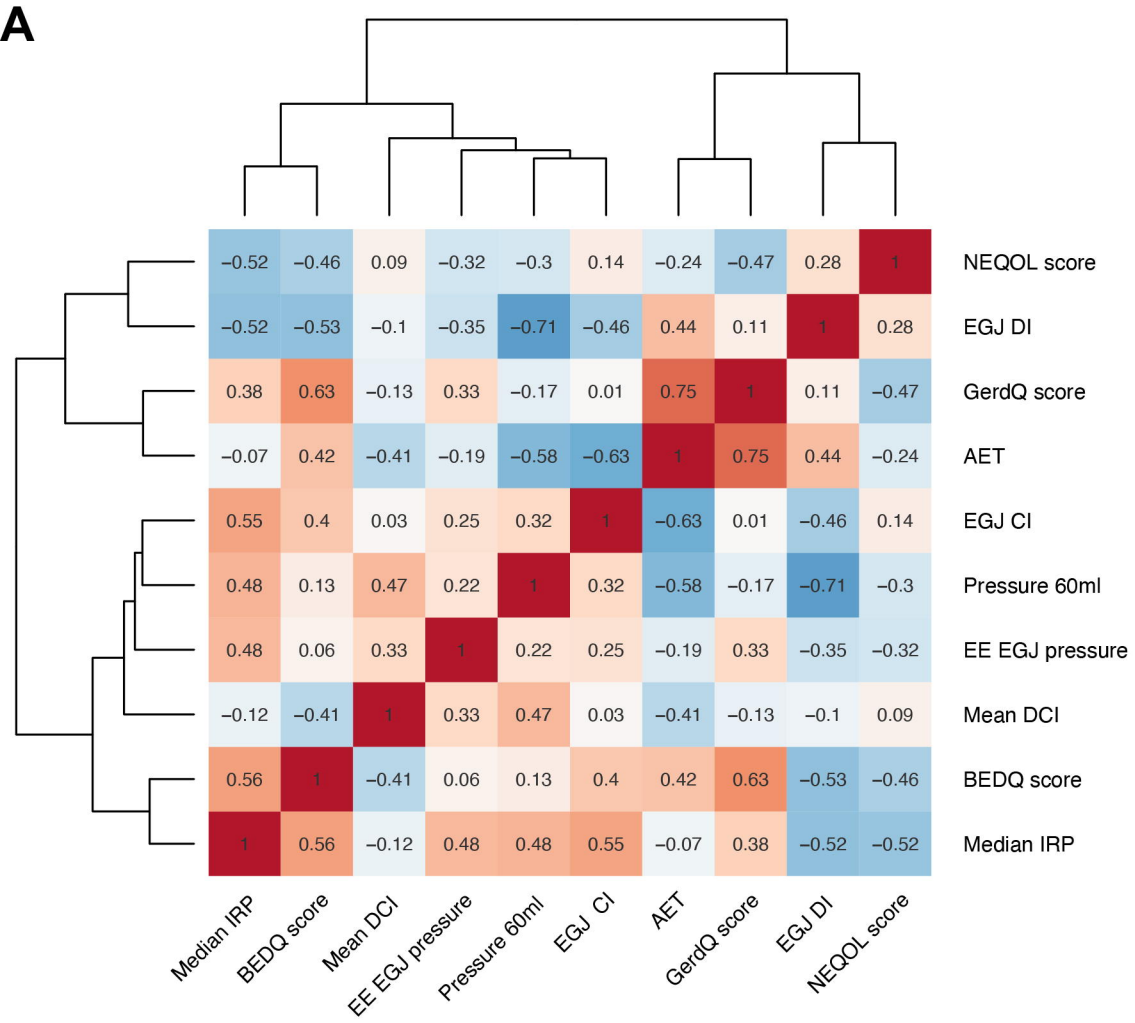

A
HC
UMAP 2
UMAP 1
Differentiation Score
0
1
GERD
SSc
B
Basal
Suprabasal
Superficial
Proliferating basal
Proliferating suprabasal
Proximal
Distal
0.0
0.2
0.4
0.6
0.8
1.0
Differentiation Score

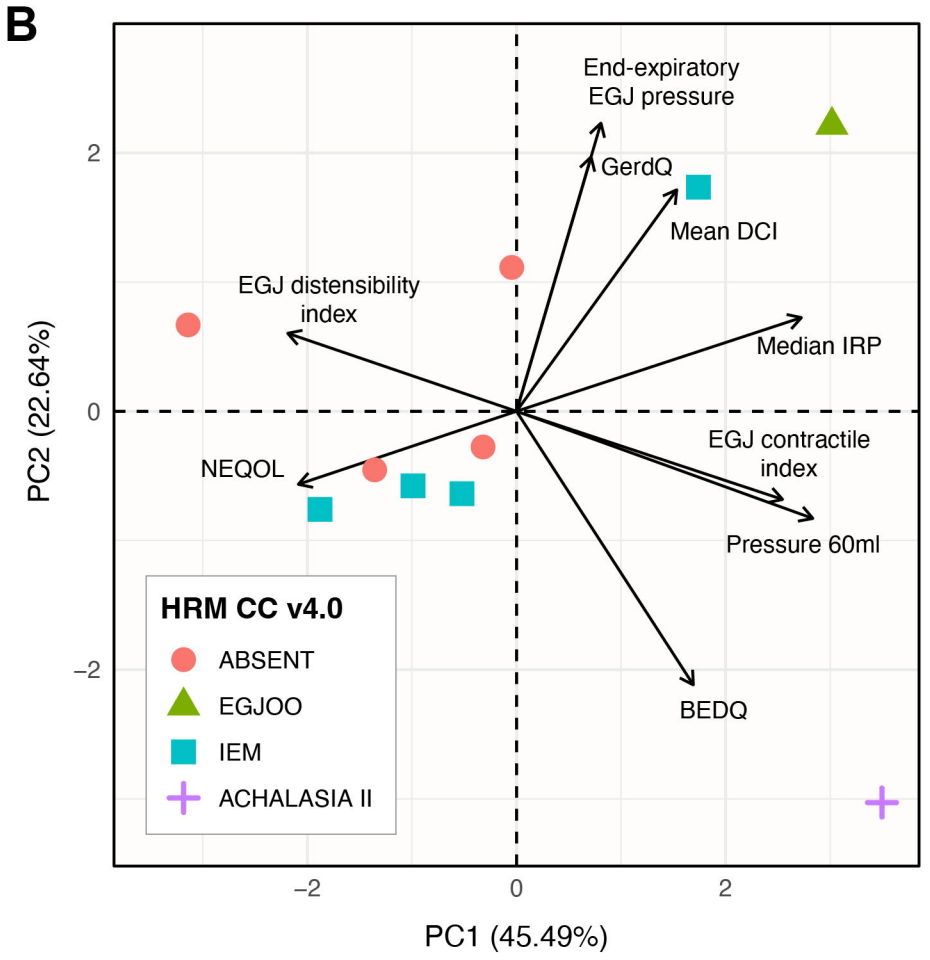

A
Basal
Proliferating Basal
Proliferating Suprabasal
Suprabasal
Superficial
Corr
Proximal
Distal
B
Basal
Prolif. basal
Prolif. suprabasal
Suprabasal
Superficial
Density
Inter-sample Gene Exp. Correlation
C
Proximal
SSc vs. HCs
GERD vs. HCs
N DEGs in Intersect
N DEGs
Intersects
D
Distal
E
DEGs vs. HCs
SSc
GERD
F
Basal
Suprabasal
Superficial
MATRISOME
MATRISOME ASSOCIATED
DEVELOPMENTAL BIOLOGY
NEUTROPHIL DEGRANULATION
KERATINIZATION
FORMATION OF THE CORNIFIED ENVELOPE
INNATE IMMUNE SYSTEM
VITAMIN D RECEPTOR PATHWAY
ECM REGULATORS
CYTOKINE SIGNALING IN IMMUNE SYSTEM
ZINC HOMEOSTASIS
NUCLEAR RECEPTORS METAPATHWAY
COPPER HOMEOSTASIS
CELLULAR RESPONSES TO STIMULI
SSc
TOLL LIKE RECEPTOR TLR1 TLR2 CASCADE
TOLL LIKE RECEPTOR CASCADES
TOLL ENDOGENOUS PATHWAY
METAPATHWAY BIOTRANSFORMATION
ANTIMICROBIAL PEPTIDES
GERD
Prox
Dist
−log10(P)
NES
G
Proximal
Epithelial Compartment
SSC vs. HCs - Log(FC)
DE Compartment
Top 25 DEGs - INNATE IMMUNE SYSTEM - Superficial
Differentiation
H
Distal
Top 25 DEGs - MATRISOME - Suprabasal

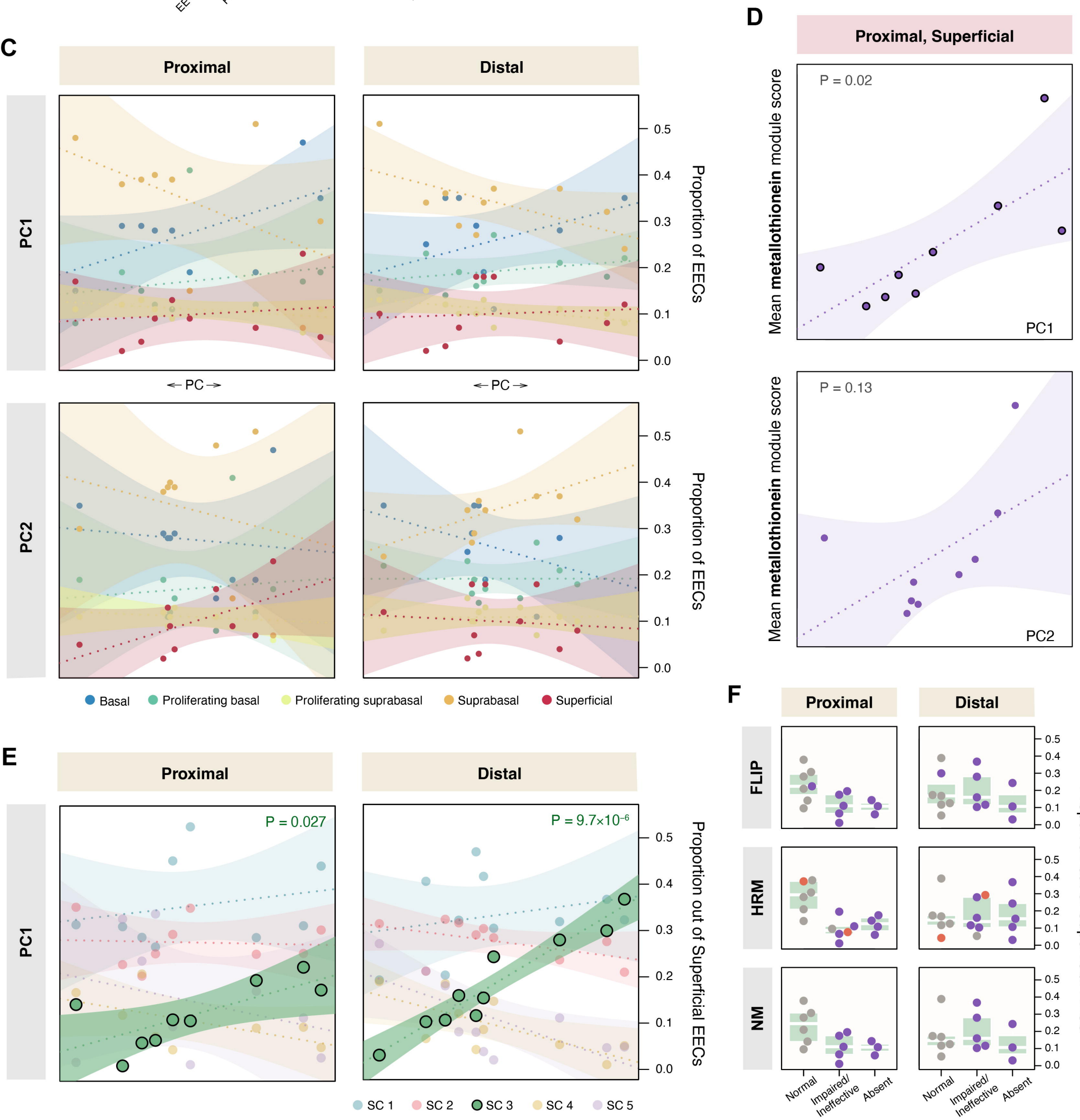

A
SSc vs. HCs DEGs
log2FC
Percent of Genes DE
Basal
Proliferating Basal
Proliferating Suprabasal
Suprabasal
Superficial
B
SSc vs. GERD DEGs
C
Superficial Proximal
r = 0.82
m = 0.87
r² = 0.64
SSc vs. HCs (log2FC)
GERD vs. HCs (log2FC)
D
Superficial Distal
r = 0.36
m = 0.44
r² = 0.26
F
Superficial Proximal
Superficial Distal
GERD
SSc
N DEGs (thousands)
G
Relative FC (m)
Proximal
Distal
SSc–HCs (log2FC) vs. GERD–HCs (log2FC)
H
r²
E
Condition
HC
GERD
SSc
Location
Proximal
Distal

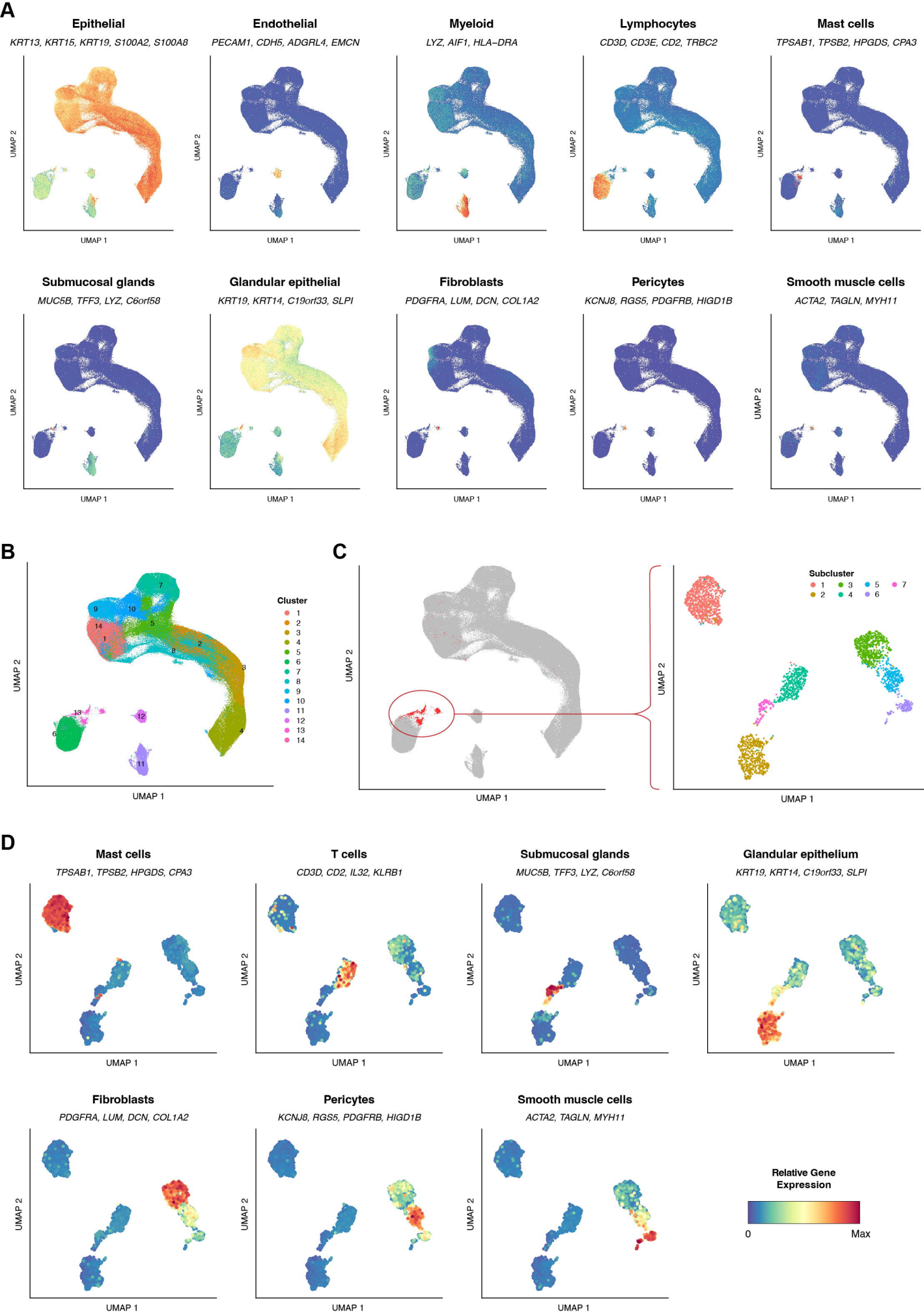

A
UMAP_2
UMAP_1
1
2
3
4
5
B
FLG
CTSV
MT1H
GJB6
FGFBP1
1 2 3 4 5
C
Differentiation Score
1.0
0.8
0.6
0.4
0.2
1 2 3 4 5
D
Proportion
0.5
0.4
0.3
0.2
0.1
0.0
*
**
*
Condition
HC
GERD
SSc
1 2 3 4 5
E
Proximal , GERD & HC
Metallothioneins
Proximal , SSc
Metallothioneins
UMAP 2
UMAP 1
Relative Expression
Max
Min
F
1 2 3 4 5
Relative Expression Metallothioneins
Condition
HC
GERD
SSc

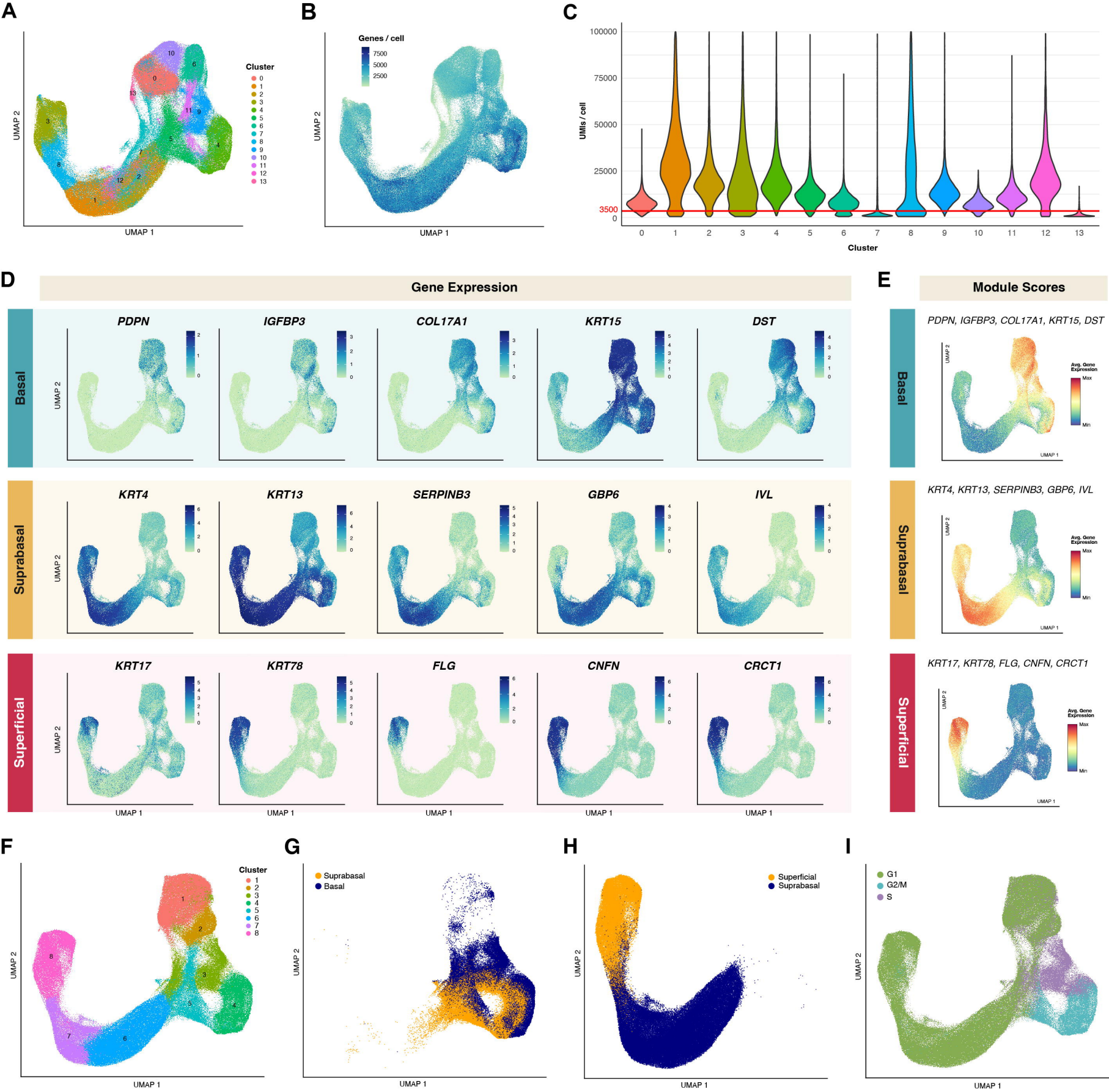

A
HCs
GERD
SSc
1
2
3
4
5
IRF1
IRF1 Target Genes
Log(Counts)
Avg. Expr.
UMAP 1
UMAP 2
B
MYC
MYC Target Genes
C
E2F4
E2F4 Target Genes
D
NFE2L2
NFE2L2 Target Genes

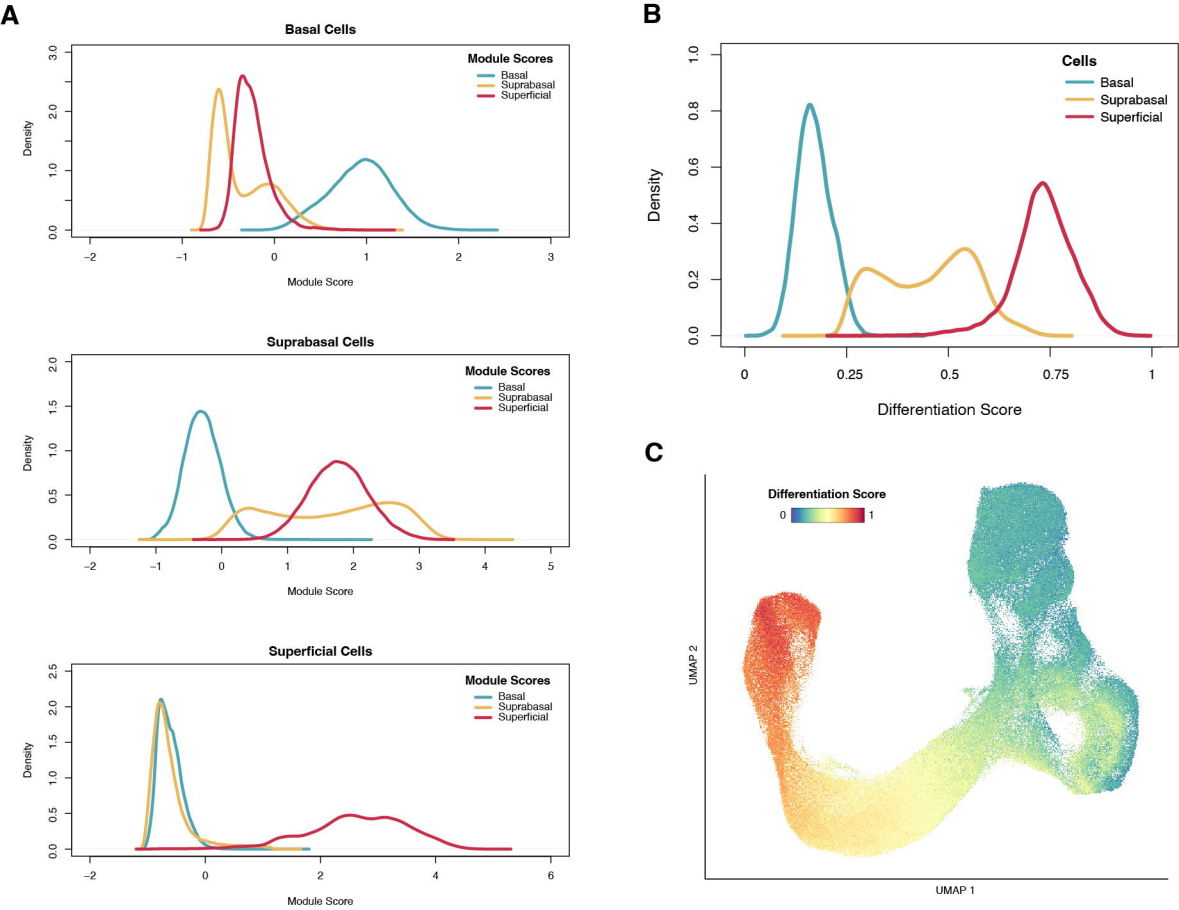

A
NEQOL score
EGJ DI
GerdQ score
AET
EGJ CI
Pressure 60ml
EE EGJ pressure
Mean DCI
BEDQ score
Median IRP
Median IRP
BEDQ score
Mean DCI
EE EGJ pressure
Pressure 60ml
EGJ CI
AET
GerdQ score
EGJ DI
NEQOL score

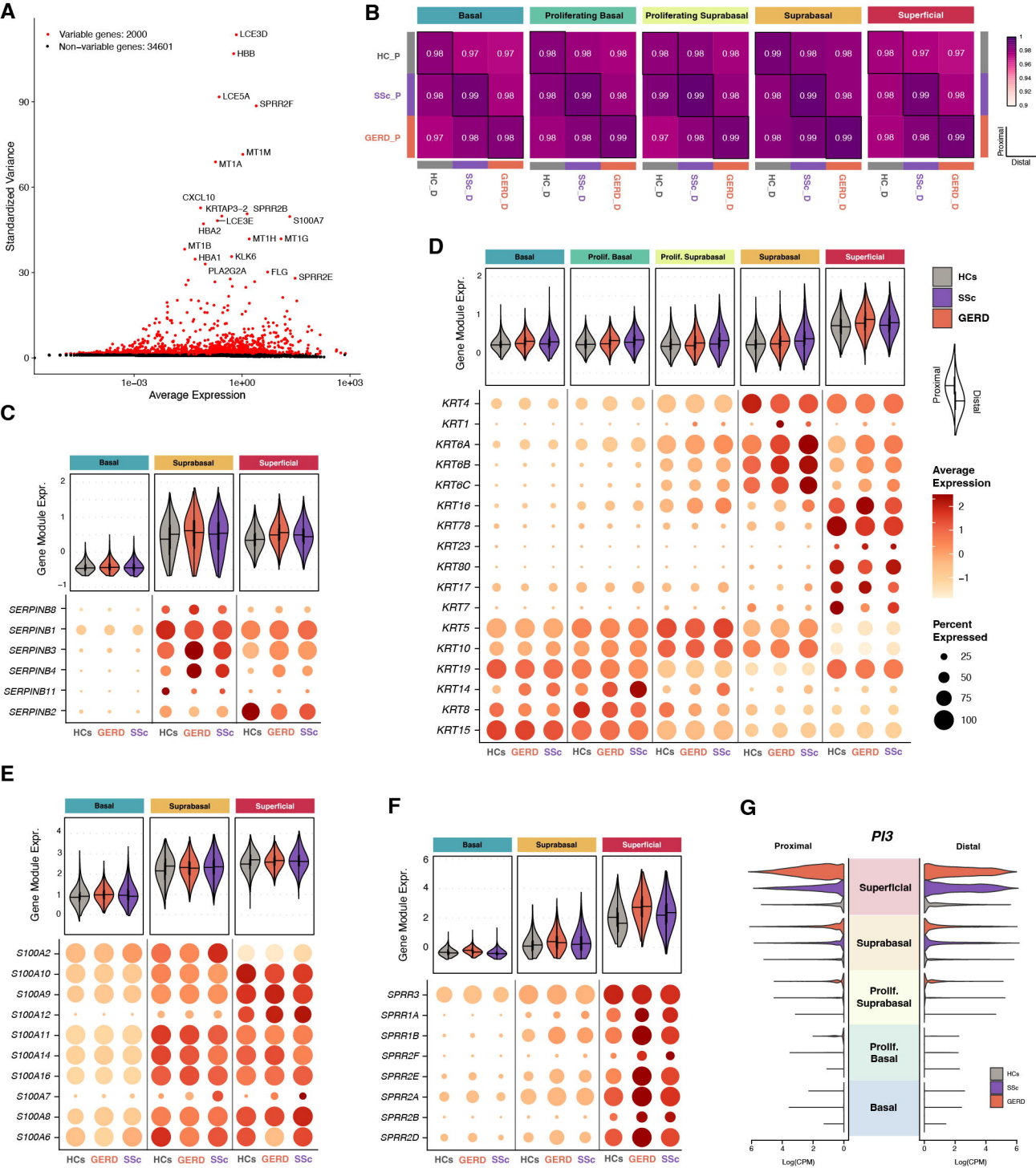

B
End-expiratory EGJ pressure
GerdQ
Mean DCI
EGJ distensibility index
Median IRP
EGJ contractile index
NEQOL
Pressure 60ml
BEDQ
HRM CC v4.0
ABSENT
EGJOO
IEM
ACHALASIA II
PC2 (22.64%)
PC1 (45.49%)

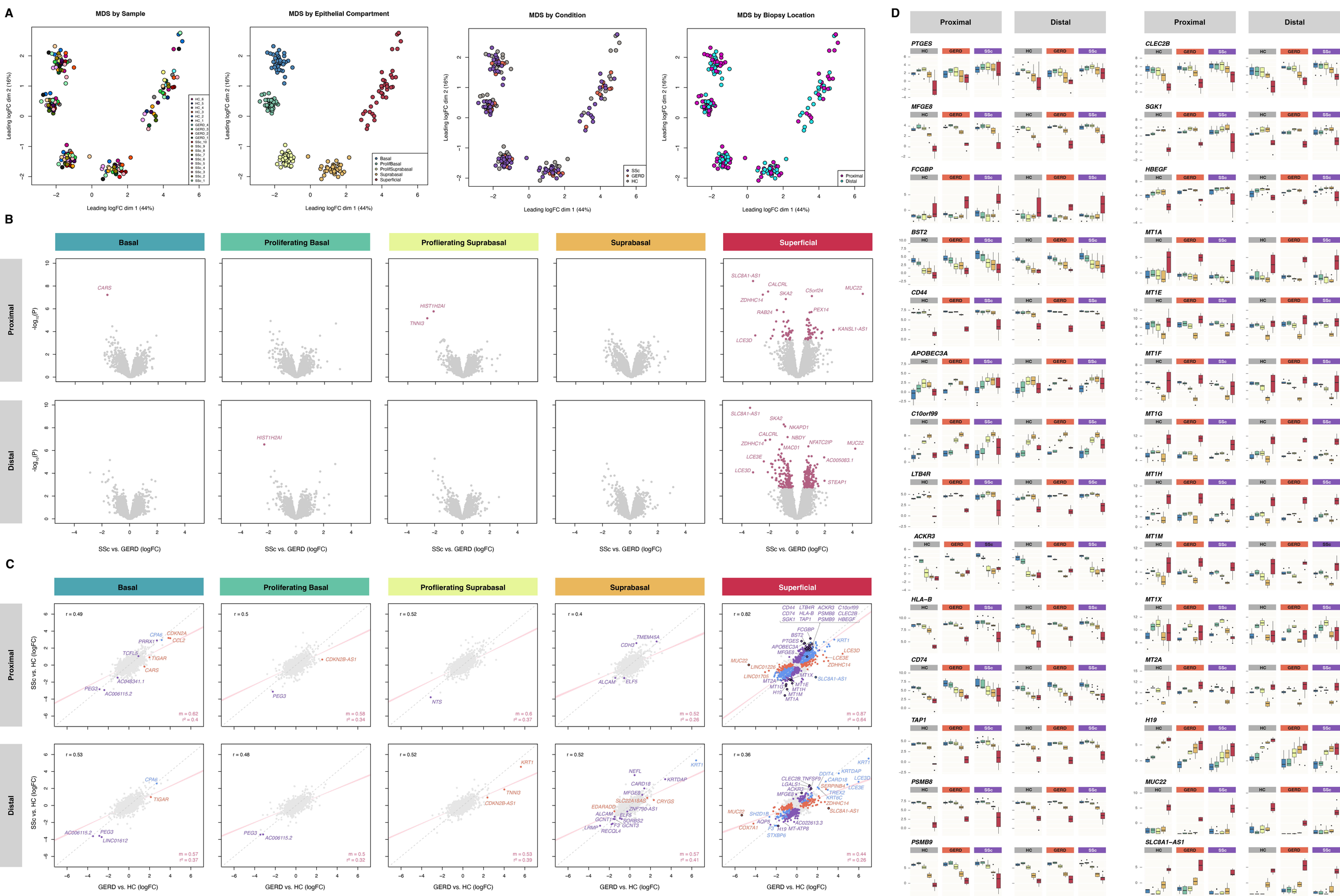

C
Proximal
Distal
PC1
PC2
← PC →
Proportion of EECs
Basal
Proliferating basal
Proliferating suprabasal
Suprabasal
Superficial
D
Proximal, Superficial
P = 0.02
Mean metallothionein module score
PC1
P = 0.13
PC2
E
P = 0.027
P = 9.7×10⁻⁶
Proportion out of Superficial EECs
SC 1
SC 2
SC 3
SC 4
SC 5
F
FLIP
HRM
NM
Normal
Impaired/ Ineffective
Absent

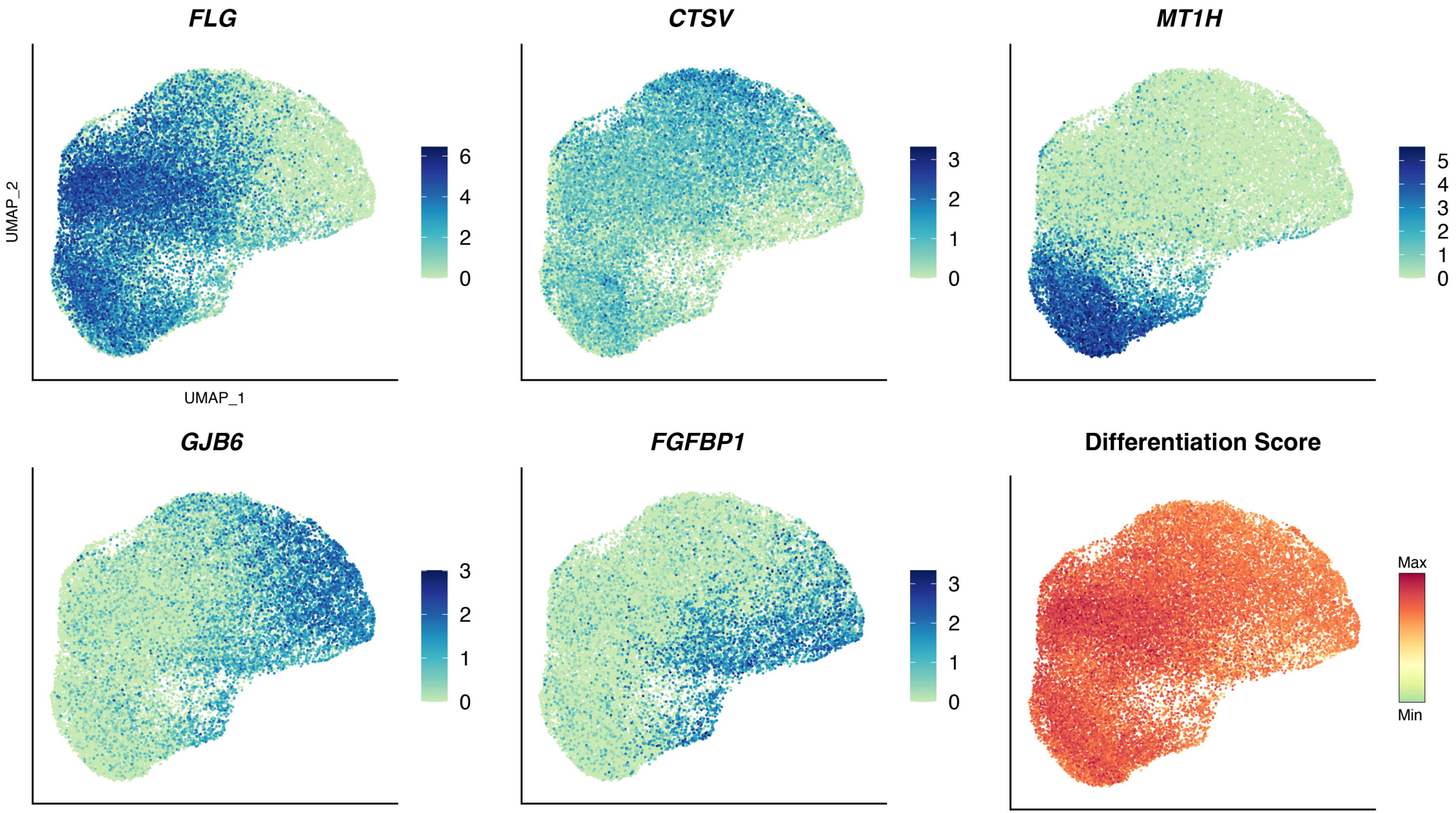

A
Epithelial
KRT13, KRT15, KRT19, S100A2, S100A8
Endothelial
PECAM1, CDH5, ADGRL4, EMCN
Myeloid
LYZ, AIF1, HLA–DRA
Lymphocytes
CD3D, CD3E, CD2, TRBC2
Mast cells
TPSAB1, TPSB2, HPGDS, CPA3
Submucosal glands
MUC5B, TFF3, LYZ, C6orf58
Glandular epithelial
KRT19, KRT14, C19orf33, SLPI
Fibroblasts
PDGFRA, 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

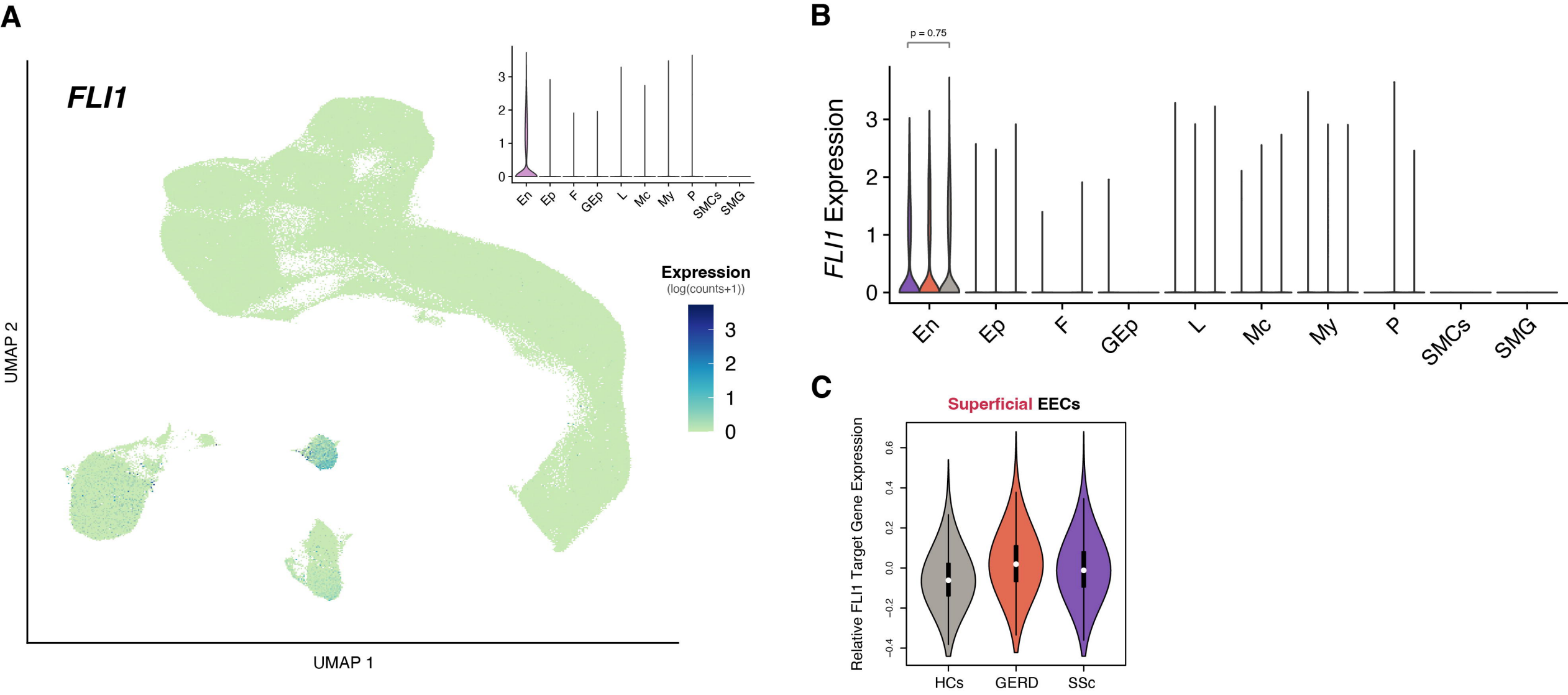

A
Cluster
UMAP 2
UMAP 1
B
Genes / cell
7500
5000
2500
UMAP 2
UMAP 1
C
UMIs / cell
100000
75000
50000
25000
3500
0
Cluster
D
Gene Expression
Basal
PDPN
IGFBP3
COL17A1
KRT15
DST
Suprabasal
KRT4
KRT13
SERPINB3
GBP6
IVL
Superficial
KRT17
KRT78
FLG
CNFN
CRCT1
UMAP 2
UMAP 1
E
Module Scores
Basal
PDPN, IGFBP3, COL17A1, KRT15, DST
Suprabasal
KRT4, KRT13, SERPINB3, GBP6, IVL
Superficial
KRT17, KRT78, FLG, CNFN, CRCT1
Avg. Gene Expression
Max
Min
F
Cluster
G
Suprabasal
Basal
H
Superficial
Suprabasal
I
G1
G2/M
S

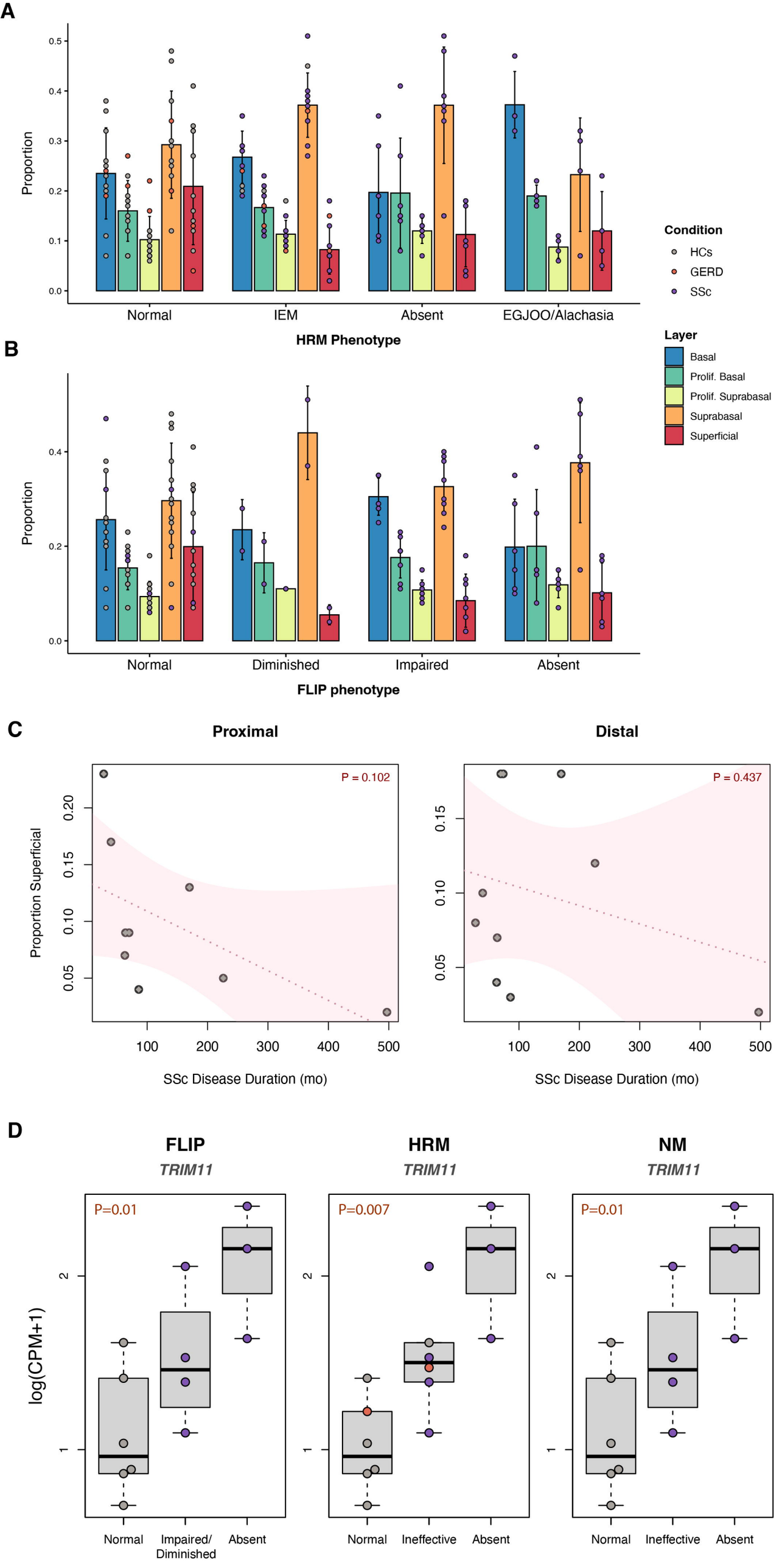

A
Basal Cells
Module Scores
Basal
Suprabasal
Superficial
Density
Module Score
Suprabasal Cells
Module Scores
Basal
Suprabasal
Superficial
Density
Module Score
Superficial Cells
Module Scores
Basal
Suprabasal
Superficial
Density
Module Score
B
Cells
Basal
Suprabasal
Superficial
Density
Differentiation Score
C
Differentiation Score
0
1
UMAP 2
UMAP 1


A
Variable genes: 2000
Non-variable genes: 34601
LCE3D
HBB
LCE5A
SPRR2F
MT1M
MT1A
CXCL10
KRTAP3-2
SPRR2B
LCE3E
S100A7
HBA2
MT1H
MT1G
MT1B
HBA1
KLK6
PLA2G2A
FLG
SPRR2E
Standardized Variance
Average Expression
B
Basal
Proliferating Basal
Proliferating Suprabasal
Suprabasal
Superficial
HC_P
SSc_P
GERD_P
HC_D
SSc_D
GERD_D
Proximal
Distal
C
Basal
Suprabasal
Superficial
Gene Module Expr.
SERPINB8
SERPINB1
SERPINB3
SERPINB4
SERPINB11
SERPINB2
HCs
GERD
SSc
D
Basal
Prolif. Basal
Prolif. Suprabasal
Suprabasal
Superficial
Gene Module Expr.
HCs
SSc
GERD
Proximal
Distal
KRT4
KRT1
KRT6A
KRT6B
KRT6C
KRT16
KRT78
KRT23
KRT80
KRT17
KRT7
KRT5
KRT10
KRT19
KRT14
KRT8
KRT15
Average Expression
Percent Expressed
E
Basal
Suprabasal
Superficial
Gene Module Expr.
S100A2
S100A10
S100A9
S100A12
S100A11
S100A14
S100A16
S100A7
S100A8
S100A6
HCs
GERD
SSc
F
Basal
Suprabasal
Superficial
Gene Module Expr.
SPRR3
SPRR1A
SPRR1B
SPRR2F
SPRR2E
SPRR2A
SPRR2B
SPRR2D
HCs
GERD
SSc
G
PI3
Proximal
Distal
Superficial
Suprabasal
Prolif. Suprabasal
Prolif. Basal
Basal
HCs
SSc
GERD
Log(CPM)


A
MDS by Sample
MDS by Epithelial Compartment
MDS by Condition
MDS by Biopsy Location
Leading logFC dim 1 (44%)
Leading logFC dim 2 (16%)
Basal
ProlifBasal
ProlifSuprabasal
Suprabasal
Superficial
SSc
GERD
HC
Proximal
Distal
B
Basal
Proliferating Basal
Proflierating Suprabasal
Suprabasal
Superficial
Proximal
Distal
-log10(P)
SSc vs. GERD (logFC)
CARS
HIST1H2AI
TNNI3
SLC8A1-AS1
CALCRL
SKA2
C5orf24
MUC22
ZDHHC14
RAB24
PEX14
KANSL1-AS1
LCE3D
NKAPD1
NBDY
NFATC2IP
MAC01
AC005083.1
STEAP1
C
Basal
Proliferating Basal
Proflierating Suprabasal
Suprabasal
Superficial
Proximal
Distal
SSc vs. HC (logFC)
GERD vs. HC (logFC)
r = 0.49
m = 0.62
r² = 0.4
CPA6
CDKN2A
PRRX1
CCL2
TCFL5
TIGAR
CARS
AC048341.1
PEG3
AC006115.2
r = 0.5
m = 0.58
r² = 0.34
CDKN2B-AS1
r = 0.52
m = 0.6
r² = 0.37
TMEM45A
CDH3
ALCAM
ELF5
NTS
r = 0.4
m = 0.52
r² = 0.26
r = 0.82
m = 0.87
r² = 0.64
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
H19
MT1H
MT1M
MT1A
r = 0.53
m = 0.57
r² = 0.37
CPA6
TIGAR
AC006115.2
PEG3
LINC01612
r = 0.48
m = 0.5
r² = 0.32
PEG3
AC006115.2
r = 0.52
m = 0.53
r² = 0.39
KRT1
TNNI3
CDKN2B-AS1
r = 0.52
m = 0.57
r² = 0.41
NEFL
KRTDAP
CARD18
MFGE8
SLC22A18AS
CRYGS
EDARADD
ZNF790-AS1
ALCAM
ELF5
GCNT1
SORBS2
LRMP
F3
GCNT3
RECQL4
r = 0.36
m = 0.44
r² = 0.26
KRT1
DDIT4
KRTDAP
CLEC2B
TNFSF9
CARD18
LCE3D
LGALS1
SERPINB4
LCE3E
ACKR3
TREX2
MFGE8
KRT6C
ZDHHC14
SLC8A1-AS1
MUC22
SH2D1B
AQP5
AC022613.3
COX7A1
F3
H19
MT-ATP8
STXBP6
D
Proximal
Distal
PTGES
MFGE8
FCGBP
BST2
CD44
APOBEC3A
C10orf99
LTB4R
ACKR3
HLA–B
CD74
TAP1
PSMB8
PSMB9
CLEC2B
SGK1
HBEGF
MT1A
MT1E
MT1F
MT1G
MT1H
MT1M
MT1X
MT2A
H19
MUC22
SLC8A1–AS1
HC
GERD
SSc


FLG
CTSV
MT1H
GJB6
FGFBP1
Differentiation Score
UMAP_2
UMAP_1
6
4
2
0
3
2
1
0
5
4
3
2
1
0
3
2
1
0
3
2
1
0
Max
Min


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
Layer
Basal
Prolif. Basal
Prolif. Suprabasal
Suprabasal
Superficial
B
Proportion
0.0
0.2
0.4
Normal
Diminished
Impaired
Absent
FLIP phenotype
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