Spatially defined multicellular functional units in colorectal cancer revealed from single cell and spatial transcriptomics

Inbal Avraham-Davidi[1,20]*, Simon Mages[1,2,20], Johanna Klughammer[1,2,20], Noa Moriel[3], Shinya Imada[4], Matan Hofree[1,] Evan Murray[5], Jonathan Chen[5,6,7], Karin Pelka[5,6,8,9], Arnav Mehta[5,6,10], Genevieve M. Boland[5,11], Toni Delorey[1], Leah Caplan[1], Danielle Dionne[1], Robert Strasser[2], Jana Lalakova[12], Anezka Niesnerova[12], Hao Xu[12], Morgane Rouault[12], Itay Tirosh[13], Nir Hacohen[5,6], Fei Chen[5,14], Omer Yilmaz[4,15], Jatin Roper[4,16,*], Orit Rozenblatt-Rosen[1,17,*], Mor Nitzan[3,18,19*], Aviv Regev[1,4,17*]

[1]Klarman Cell Observatory, Broad Institute of MIT and Harvard, Cambridge, MA, USA

[2]Gene Center and Department of Biochemistry, Ludwig-Maximilians-Universität München, Munich, Germany

[3]School of Computer Science and Engineering, The Hebrew University of Jerusalem, Jerusalem, Israel

[4]Departmen of Biology at MIT, Koch Institute for Integrative Cancer Research at MIT, Cambridge, MA, USA

[5]Broad Institute of Massachusetts Institute of Technology and Harvard, Cambridge, MA, USA

[6]Massachusetts General Hospital (MGH) Cancer Center, Harvard Medical School (HMS), Boston, MA, USA

[7]Department of Pathology, MGH, Boston, MA, USA

[8]Current address: Gladstone-UCSF Institute of Genomic Immunology, Gladstone Institutes, CA, USA

[9]Current address: Department of Microbiology and Immunology, UCSF, CA, USA

[10]Dana Farber Cancer Institute, Boston, MA, USA

24 [11]Department of Surgery, MGH, Boston, MA, USA

25 [12]10xGenomics, Solnavägen 3H, Stockholm, Sweden

26 [13]Department of Molecular Cell Biology, Weizmann Institute of Science, Rehovot, Israel

27 [14]Harvard Stem Cell and Regenerative Biology, Cambridge MA, USA

28 [15]Department of Pathology, Massachusetts General Hospital, Boston, MA, USA

29 [16]Current address: Department of Pharmacology and Cancer Biology and Department of Medicine,

30 Division of Gastroenterology, Duke University, Durham, NC, USA

31 [17]Current address: Genentech, 1 DNA Way, South San Francisco, CA, USA

32 [18]Racah Institute of Physics and Faculty of Medicine, The Hebrew University of Jerusalem,

33 Jerusalem, Israel

34 [19]Faculty of Medicine, The Hebrew University of Jerusalem, Jerusalem, Israel

35 [20]These authors contributed equally

36 *Correspondence: inbalavr@gmail.com (I.A.-D.), jatin.roper@duke.edu (J.R.),

37 orit.r.rosen@gmail.com (O.R.-R.), mor.nitzan@mail.huji.ac.il (M.N.), aviv.regev.sc@gmail.com

38 (A.R.)

Abstract

While advances in single cell genomics have helped to chart the cellular components of tumor ecosystems, it has been more challenging to characterize their specific spatial organization and functional interactions. Here, we combine single cell RNA-seq, spatial transcriptomics by Slide-seq, and *in situ* multiplex RNA analysis, to create a detailed spatial map of healthy and dysplastic colon cellular ecosystems and their association with disease progression. We profiled inducible genetic CRC mouse models that recapitulate key features of human CRC, assigned cell types and epithelial expression programs to spatial tissue locations in tumors, and computationally used them to identify the regional features spanning different cells in the same spatial niche. We find that tumors were organized in cellular neighborhoods, each with a distinct composition of cell subtypes, expression programs, and local cellular interactions. Comparing to scRNA-seq and Slide-seq data from human CRC, we find that both cell composition and layout features were conserved between the species, with mouse neighborhoods correlating with malignancy and clinical outcome in human patient tumors, highlighting the relevance of our findings to human disease. Our work offers a comprehensive framework that is applicable across various tissues, tumors, and disease conditions, with tools for the extrapolation of findings from experimental mouse models to human diseases.

INTRODUCTION

The spatial organization of diverse cells in the tumor ecosystem impacts and drives interactions between malignant cells and neighboring immune and stromal cells, either promoting or suppressing tumor growth[1]. Recent studies have shown that systematic understanding of the spatial organization of tumors can shed light on disease progression and response to therapy, with specific features correlated with tumor subtypes[2–4], cancer prognosis[5–7], or response to treatment[8,9].

However, genome-scale, high-resolution dissection of the spatial organization of tumors and its functional implications remains challenging, largely due to technical limitations. Methods such as fluorescent *in situ* hybridization (FISH) and immunohistochemistry can only measure a handful of pre-selected transcripts or proteins, whereas single cell RNA-seq (scRNA-seq) does not directly capture spatial relations. Recent advances in spatial genomics and proteomics allow multiplexed or genome-scale measurements *in situ*[10–17], but with a trade-off between genomic scale and spatial resolution[18]. As a result, data from different experimental methods need to be integrated for a comprehensive view of the tissue biology. Many analytical tools have been developed to integrate some crucial aspects of the data[19–23], but it can be challenging to deploy them and distill answers to specific disease biology questions. This leaves open many fundamental questions about tissue organization and collective function, including whether there are canonical functional units in tumors, what may be their organization in the tumor landscape, and what role does each play in tumor progression.

78 A case in point is colorectal cancer (CRC), where initial lesions (adenomatous polyps) progress
79 over time to carcinoma and eventually to metastatic disease. While the mutations that drive this
80 process were extensively studied[24–27], and the cellular ecosystem of CRC has now been deeply
81 charted[2,28,29], the spatial landscape is less well-characterized. In a recent study of human CRC[2], we
82 statistically associated cell profiles across tumors and showed that they map to different cellular
83 communities, reside in different locations in the tumor and reflect different tumor subtypes.
84 However, absent genome wide *in situ* measurements, these statistical inferences do not yet reflect
85 the full spatial organization of the tumor. Moreover, it is important to relate such patterns to those
86 in animal models used in mechanistic studies and as pre-clinical models, to understand the relation
87 between lab models and human tumors.
88
89 Here, we deciphered the spatial and cellular organization of colorectal cancer (CRC) by combining
90 scRNA-seq, spatial transcriptomics by Slide-seq, and *in situ* RNA multiplex analysis, using a novel
91 computational framework. We first profiled two inducible genetic mouse models of colorectal
92 cancer that recapitulate key features of human CRC, before and after tumor initiation. We
93 integrated the spatial and cell profiles to create a spatial cell map of the tumor landscape, revealing
94 dysplasia-specific cellular layout and potential physical interactions. We found that the tumor
95 landscape is organized in cell neighborhoods, each with distinct epithelial, immune, and stromal
96 cell compositions, and governed by different gene programs. Three of the cell neighborhoods are
97 associated with tumor progression, each activating different biological pathways but all active
98 simultaneously albeit in different parts of the tumor. We devised a computational framework,
99 based on the TACCO[21] method, extending it to compare single cell and spatial features of tumors
100 between species and applied it to scRNA-seq data from human CRC. Multiple features were

101 conserved between tumors in the mouse model and the human patients, and the mouse cellular
102 neighborhoods correlated with malignancy and clinical outcome (progression-free intervals (PFI)
103 and overall survival (OS)) in human patient tumors. Our work provides a general approach that
104 can be applied to other tissues, tumors and disease conditions.

105

106 **RESULTS**

107 **A cell atlas of genetic models of colorectal cancer**

108 To chart the cell ecosystem of CRC and how it changes during tumor progression, we studied two
109 genetic mouse models of CRC, one with inactivation of *Apc* and another in which *Apc* inactivation
110 is accompanied by an oncogenic $Kras^{G12D/+}$ mutation and inactivation of *Trp53*[30–32] (**Figure 1A**).
111 In the AV model, $Apc^{fl/fl}Villin^{creERT2}$ mice are injected with 4-hydroxytamoxifen to the submucosal
112 layer of the colon, inducing the deletion of *Apc* specifically in epithelial cells within the injection
113 site[30]. In the AKPV model ($Apc^{fl/fl}$; $LSL\text{-}Kras^{G12D}$; $Trp53^{fl/fl}$;$Rosa^{26LSL\text{-}tdTomato/+}$; $Villin^{CreERT2}$ mice,
114 **Methods**), 4-hydroxytamoxifen injection also induces an oncogenic $Kras^{G12D/+}$ mutation and then
115 inactivation of *Trp53*. In both cases, 4-hydroxytamoxifen injection leads to the formation of local
116 lesions that resemble human dysplastic lesions[30].

117

118 We first generated a single-cell atlas of the models consisting of 48,115 high quality scRNA-seq
119 profiles from normal colon, AV (3 weeks after 4-hydroxytamoxifen induction) and AKPV (3 and
120 9 weeks after induction) tissues. We captured a diverse cell census (**Figure 1B,C**, **Methods**), with
121 35 clusters annotated *post hoc* by the expression of known marker genes (**Figure 1B**, **Figure**
122 **S1A,B**, **Methods**), across epithelial, immune and stromal cell compartments (**Figure S1C**). For

123 validation purposes, we also generated multiplex *in situ* RNA profiles (with the Cartana[33] method)
124 in 6 sections each from normal and AV conditions, using a panel of 66-180 marker genes chosen
125 to best represent the cell types and programs found in the tissues (**Figure S1D, Methods**).
126

127 **Tumorigenic genotypes cause shifts in the composition of epithelial cell populations and their**
128 **microenvironment in AV and AKPV lesions**

129 Dysplastic lesions exhibited shifts in proportions of immune and stromal cells, including changes
130 in subsets pre-existing in normal tissue, as well as infiltration of new cell subsets (**Figure 1C,D**
131 and **Figure S1D,E,F**). This resulted in both increase in cells of existing populations (*e.g.*, some T
132 cell subsets (TNK05 (GdT/Il17+), TNK06 (Treg)) and emergence of new dysplasia-associated
133 cells (*e.g.*, granulocytes (Gran01, Gran02) and monocytes (Mono02, Mono03)) mirroring
134 observations in human CRC[2], breast cancer[34], and non-small cell lung cancer[35] (**Figure 1C,D** and
135 **Figure S1D,E** and **Figure S2 A-C**). We validated these patterns using multiplex *in situ* RNA
136 analysis (**Figure 1E**, **Figure S1F-H** and **Supplementary item 1 and 2**). Infiltration is likely to
137 underlie many of these changes as many of the increasing cell subsets (granulocytes, monocytes,
138 mast cells) expressed genes, such as Sell and Ccr2, indicating tissue recruitment, and as the cells
139 dramatically increase in proportion despite negligible signals of proliferation programs.

140 Two of four monocyte subsets, Mono02 and Mono03, were unique dysplasia-associated cells
141 (**Figure S2D-F**) and were respectively enriched for general inflammatory response genes
142 (FDR=5.7 10^{-30}, two-sided Fisher's exact test in GO term enrichment) and interferon beta and
143 gamma response genes (FDR=3.5 10^{-11}, 1.0 10^{-13}). T cell subsets showed the expected diversity
144 across nine subsets (**Figure S2G-I**)[36], with a significant decrease (out of all T cells) in TNK01
145 (GdT/Cd8) in the dysplastic microenvironment and an increase in TNK05 (GdT/Il17+) (**Figure**

146 **S2J**). This is consistent with the T cell composition in tumors from mismatch repair proficient
147 (MMRp) CRC patients (**Figure S2K**). RNA velocity analysis[37] of T cells from normal and AV
148 tissue (**Figure S3A**) showed a change in inferred cellular relationships with TNK03 (naïve T) and
149 TNK02 (Th1/Th17) preceding TNK06 (Treg), consistent with the promotion of an
150 immunosuppressive microenvironment, and TNK08 (proliferating T) also preceding TNK02
151 (Th1/Th17), TNK04 (GdT/Cd8) and TNK05 (GdT/Il17+) populations.
152
153 Within the five subsets of stromal cells (including vascular endothelial and lymphatic endothelial
154 cells and three fibroblast subsets, **Figure S2L-N** and **Methods**), Endo01 (vascular) were enriched
155 in dysplastic lesions compared to normal colon (FDR=1.8 10^{-3}, two-sided Welch's t test on CLR
156 transformed compositions; **Figure S2O**). Angiogenesis-related pathways, such as angiogenesis
157 (FDR=1.0 10^{-11}) and positive regulation of angiogenesis (FDR=3.4 10^{-4}) as well glycolytic process
158 (FDR=1.4 10^{-4}) were enriched in Endo01 (vascular) from lesions compared to normal colon (**Table**
159 **S1**). This is in line with vascular adaptation to the tumor's growing needs for nutrients and
160 oxygen[38] and with the increased expression of the vascular growth factor *Vegf-A* in both monocytes
161 and macrophages (**Figure S2P**).

162 **Cell-intrinsic expression shifts in different sub-lineages in the dysplastic epithelium**

163 Epithelial cells showed dramatic cell-intrinsic changes between normal tissues and either AV or
164 AKPV lesions, such that the cell profiles of dysplastic epithelial cells in both models were highly
165 distinct from normal epithelial cells (and similar to each other) (**Figure 1C**, **Figure S4A** and
166 **Methods**). Epithelial cell profiles from normal mice (41% of cells) separated from most of those
167 from AV and AKPV models (59% of cells) (**Figure 2A** and **Figure S4A,B**), suggesting a common
168 shift in all dysplastic cells from the normal state. Notably, 11% of epithelial cells from AV/AKPV

169 mice were classified as non-dysplastic healthy cells, indicating that normal, non-dysplastic, cells
170 may be present in or adjacent to the lesion microenvironment (**Figure 2B** and **Figure S4A**),
171 although we cannot firmly rule out adjacent tissue contamination. We annotated two cell clusters
172 – Epi01 (dysplastic stem-like) and Epi05 (dysplastic secretory-like) – as dysplastic, due to their
173 virtually exclusive presence in AV and AKPV models and because they expressed high levels of
174 *Apc* target genes (*e.g.*, *Axin2, Ascl2, Myc, Ccnd1, Lgr5*) and were enriched in tdTomato^{+} cells from
175 AKPVT mice (**Figure 2C** and **Figure S4C,D**) while also spanning the enterocyte to secretory
176 continuum with healthy cells (**Figure S4B**, PC2).
177
178 Interestingly, Epi05 (dysplastic secretory-like) had distinguishing markers (*e.g.*, *Ccl9, Mmp7,*
179 *Ifitm3*) from their counterparts in normal tissue, Epi04 (secretory) (**Figure 2A-C** and **Figure**
180 **S4A,C,E**). *Mmp7* and *Ifitm3* are known to promote metastasis in human CRC[39,40], and *Ccl9*
181 expression by epithelial cells promotes tumor invasion through recruitment of Ccr1^{+} myeloid cells
182 to the tumor's invasive front in a mouse model of CRC[41]. Notably, *Ccr1* is expressed by newly
183 recruited monocytes, macrophages and granulocytes in our model, suggesting a potential
184 mechanism for tumor infiltration and invasion (**Figure S4E**). Thus, dysplastic secretory epithelial
185 cells may perform additional functions that differ from those of their healthy counterparts.
186
187 RNA velocity[37] analysis of the epithelial cell compartment predicted that in the dysplastic
188 epithelium (**Figure S3B, right**) a proliferating stem-cell like dysplastic subset is a direct source to
189 both a massively expanded and heterogeneous non-proliferating population of dysplastic stem-like
190 cells (expressing WNT signaling and angiogenesis programs) and to dysplastic cells of different
191 'differentiation states' (MHCII expressing stem/progenitors leading to enterocytes and secretory-

192 like cells) (**Figure S3B,C**). Conversely, in the normal epithelium (**Figure S3B**, left), the 'root' is
193 placed in a much smaller population of proliferating intestinal stem cells, and the stem cell
194 compartment is overall much more modest. The proliferation and differentiation path of previously
195 identified normal epithelial cells remains intact (*i.e.*, reminiscent of the one in normal samples)
196 even in the dysplastic lesions (**Figure S3B**). While our RNA velocity analysis provides insights
197 into potential cellular trajectories, further experimental validation is required to confirm these
198 findings.
199
200 **Expression programs for stem-like functions, Wnt signaling, angiogenesis and inflammation**
201 **are activated in dysplastic epithelial cells**
202 Both normal and dysplastic epithelial cells varied along a continuum, as expected and previously
203 observed in the ongoing differentiation in the colon epithelium[2,42–44] and our RNA Velocity
204 analysis (**Figure S3**). Using non-negative matrix factorization (iNMF from LIGER[45], **Methods**),
205 we recovered 20 expression programs spanning the different epithelial functions, and annotated
206 them by Gene Ontology terms enriched in their top 100 weighted genes (**Figure 2D,E**, **Figure**
207 **S3C, S4F-L, Methods**).
208
209 The programs enriched in different dysplastic cells highlighted key processes that play a role in
210 tumor promotion, including stem cell programs, Wnt signaling, angiogenesis, and inflammation
211 and innate immunity, including interferon alpha, beta and gamma pathways (**Figure 2D,E, Figure**
212 **S3C**). In particular, the stem cell program (#16) detected in some cells across all conditions, was
213 enriched in dysplastic samples (FDR=5.6 10^{-10}, two-sided Welch's t-test on CLR transformed
214 compositions), reminiscent of a recently described population in human[28,29]. Comparing cells from

215 dysplastic and normal samples that express the stem cell program, the dysplastic cells had distinct
216 expression profiles with induction of negative regulators of Wnt signaling (FDR=4.5 10^{-5}, two-
217 sided Fisher's exact test in GO term enrichment, *e.g.*, *Notum*, Wnt inhibitory factor 1 (*Wif1*) and
218 *Nkd1*) and genes that are related to cellular response to interferon-gamma (FDR=1.7 10^{-6}, *e.g.*,
219 *Ccl9, Ccl6*) and immune system process (FDR=6.4 10^{-4}, *e.g.*, Ifitm1 and Ifitm3) (**Figure 2E,F**).
220 This is consistent with recent studies showing that *Apc*-mutant stem cells secrete negative
221 regulators of Wnt signaling to induce the differentiation of the WT stem cells in their proximity,
222 thereby outcompeting them and promoting tumor formation[46,47]. Thus, stem cells from dysplastic
223 lesions may have non-canonical function and regulation. In addition, the programs for Wnt
224 signaling (expressing both positive and negative regulators; #4, FDR=2.8 10^{-6}), angiogenesis (#14,
225 FDR=1.2 10^{-9}), inflammatory response (#6, FDR=1.4 10^{-6}), and innate immune response and
226 interferon response (#7, FDR=1.2 10^{-2}) were all predominantly expressed or enriched in
227 AV/AKPV epithelium (all with two-sided Welch's t-test on CLR transformed compositions,
228 **Figure 2E** and **Figure S4I-M**). These results are consistent with the known role of the Wnt
229 signaling pathway in CRC, and of angiogenesis, response to hypoxia and inflammation in tumor
230 progression[48–50].
231

232 **Malignant-like tissue programs and composition are conserved between mouse and human**
233 **tumors**

234 To evaluate the relevance of our findings to human colorectal cancer, we compared them to a
235 scRNA-seq atlas we recently generated from tumor and adjacent normal tissue from 62 patients
236 with either MMRp or MMRd CRC[2]. We compared mouse and human tumors in terms of their
237 epithelial expression programs, cellular composition, and cell associations in multicellular hubs[2].

238 While our mouse model should more closely resemble MMRp tumors, we compared to both
239 classes separately and together to identify any shared features.

240 To assess the similarity between mouse and human programs we controlled for overall cross-
241 species and batch differences by normalizing program-specific expression profiles with species-
242 specific background profiles (**Methods**). The Pearson correlation coefficients of these normalized
243 scores between the human and mouse programs indicate some overlap in the programs defined on
244 mouse and human data (**Figure S5A**). Epithelial cells from human and mouse tumors expressed
245 many programs highly correlated between the species (**Figure S5B**, **Methods**), including for cell
246 cycle, inflammation, epithelial secretory, angiogenesis, Wnt signaling, and normal colon
247 functions.

248 Co-variation in cell proportions across samples (by scRNA-seq) was also conserved between
249 human and mouse tumors, suggesting broad conservation of tumor composition. For example, in
250 both species the proportion of endothelial cells and fibroblasts correlated across samples, as did T
251 and B cell proportions in human tumors and mouse dysplastic lesions (**Figure S5C**). Moreover,
252 when we transferred epithelial program annotations from mouse to human scRNA-seq and
253 calculated their co-variation across samples in each species, programs 11 (proliferation), 14
254 (angiogenesis) and 16 (stem cells), co-varied both across dysplastic mouse tumors and across
255 human MMRp and MMRd tumors (**Figure S5D** and **Methods**), suggesting a conserved dysplastic
256 tissue architecture.

257

258 **Integrated spatial and single-cell atlas of mouse CRC tumors**

259 To comprehensively decipher the distribution of cells and programs in the tumor spatial niche, we
260 next used Slide-seqV2[15] for genome-wide spatial RNA-seq at 10 μm resolution. We sectioned and
261 profiled frozen tissues from four normal colon and four AV lesions using 10 Slide-seqV2 pucks
262 (**Methods**), recovering 221,936 high quality beads (**Figure S6A-C**, **Methods**). We then integrated
263 the single cell census and spatial profiles using TACCO, which allowed us to annotate each bead
264 with compositions of discrete cell types (from epithelial, immune and stromal compartments) and
265 to further annotate the epithelial fraction of each bead with a composition of epithelial program
266 activity (**Figure 1A** "annotation").
267
268 We first used TACCO to annotate every bead in the Slide-seq data with a composition of discrete
269 cell subtypes for every puck separately, using its matching single-cell reference (normal or disease;
270 **Figure 1A**, **Methods**). To this end, TACCO iteratively solved optimal transport problems to assign
271 cell subtypes to fractions of reads of the beads. TACCO relies on unbalanced optimal transport to
272 allow for shifts in the frequency of cell subtypes in the pucks *vs*. the single-cell dataset, while using
273 the reference cellular frequencies as prior knowledge (**Figure S6E,G**). TACCO's cell type
274 mapping recapitulated the muscularis layer in its expected tissue location based on the inferred
275 cellular composition pattern (**Figure 3A, 4A and S6D**). This illustrates that TACCO's
276 compositional annotations align well with biological patterns.
277
278 Next, we used TACCO to map the epithelial gene programs (defined above), focusing on transcript
279 counts that are inferred as derived from epithelial cells. TACCO partitioned the read count matrices
280 for each puck, assigning counts to epithelial cells based on the mapped per-bead cell subtype
281 annotations (from the first step) and the expression profiles associated with each subtype

282 (**Methods**). It then summed all epithelial contributions into an epithelial-only spatial count matrix,
283 followed by optimal transport to assign epithelial program contributions to individual beads, based
284 on epithelial cell-only read signals. As for cell type mapping, the proportional contribution of the
285 programs largely recapitulated their contributions in scRNA-seq (**Figure S6F,H**).
286
287 **Altered and less ordered local cellular organization of dysplastic lesions**
288 We assessed the local cellular architecture in terms of the preferential proximity of cells of certain
289 type or expressing particular epithelial programs, within a fixed-sized neighborhood, by adapting
290 an earlier method. We defined a z-score as significance of the observed neighborship relations
291 compared to the null for neighborhoods of 20, 40 or 60 μm diameter (**Figure 3B,C** and **Figure**
292 **S7A,B**). This z-score is defined with respect to a population of random cell type annotations
293 generated by random permutations of the cell type annotations between the beads, where in our
294 case we permute fractional cell type contributions.
295 Cell proximity preferences in the normal colon tissue are consistent with the expected morphology,
296 validating our approach (**Figure 3A,B**). Epithelial cells were organized such that the differentiated
297 Epi02 (Enterocytes) are excluded from the stem cell niche (**Figure 3B** and **Figure S7C**), and
298 endothelial cells and fibroblasts were also spatially co-located in a focused region (**Figure 3B**),
299 with T cells in their vicinity (**Figure 3B**). Our multiplex *in situ* RNA analysis validated the
300 exclusion of enterocytes from the stem cell niche, as also seen in Slide-seq data (**Figure S7C,D**).
301
302 While some normal tissue features are preserved in dysplastic samples, including co-location of
303 cells of the same lineage[5,10] (**Figure 3A-C**), there were notable changes, and more disorder. Cell
304 types were more randomly distributed in AV lesion *vs*. normal tissue, reflected in lower z-scores

305 (p=1.6 10^{-37}, one-sided Mann-Whitney U test; **Figure S7E**). At short distances, all epithelial cells
306 (normal and dysplastic) were preferentially located close to cells from the same subtype (**Figure**
307 **3C**) and even to cells with similar functions: epithelial cells expressing programs associated with
308 malignant-like function (*e.g.*, program 4 (Wnt signaling), 14 (angiogenesis) and 16 (stem cells))
309 resided close to each other and were spatially distant from cells expressing programs that are
310 related to normal epithelial functions (*e.g.*, program 5 (basolateral plasma membrane), 8 (apical
311 plasma membrane) and 10 (oxidation-reduction process)), supporting a model where tumor
312 progression is structured and compartmentalized (**Figure S7F**). Immune and stromal cells were
313 generally excluded from epithelial cell neighborhoods. Granulocytes aggregated together (self-
314 proximal) (**Figure 3C**) and were relatively close to endothelial cells and dysplasia-associated
315 monocytes (Mono02, Mono03), consistent with their recruitment from the blood through the
316 vessels (**Figure 3C** and **Figure S7G**). Our multiplex *in situ* RNA analysis validated the spatial
317 enrichment of monocytes and granulocytes near vessels (**Figure S7H**).
318
319 **Epithelial regional analysis recovers canonical structures in normal colon**
320 To detect distinctive tissue regions in tumors, which lack traditional tissue references, we identified
321 cellular neighborhoods with both similar epithelial program activity and a particular composition
322 of immune and stromal cells. Specifically, we first identified "epithelial program regions" as areas
323 of distinct epithelial program activity, and then found immune or stromal cells associated with
324 each region (**Figure 1A** "annotation"). Intuitively, we defined "regions" based on both the
325 similarity in epithelial expression program activity and proximity in space. To do this, after
326 assigning epithelial programs to epithelial beads, we clustered the beads based on a weighted sum
327 of spatial proximity and expression program similarity. This results in spatially contiguous

328 annotation of beads with distinct epithelial program activity, which, together with the immune and
329 stromal cells in their proximity, compose the "region". Specifically, using TACCO, we defined
330 epithelial program regions by Leiden clustering of the weighted sum of neighborship graphs for
331 spatial bead proximity and epithelial expression program similarity, such that transcriptionally
332 similar epithelial beads on different pucks can be connected (despite "infinite" spatial distance,
333 **Methods**). We then used this single framework for region annotation across all pucks (**Figure 4A**),
334 to determine the distinctive composition of additional cell types in the same set of spatial regions
335 (**Figure 4B-D** and **Figure S8A**).
336
337 In the normal colon, the regional analysis (**Figure 4A**, bottom) robustly recovered the expected
338 spatial organization of the healthy colon across five regions and their cellular composition and
339 sublayers (**Figure 1A**), from luminal/apical to basal. Four regions recovered by TACCO
340 corresponded to different layers of the mucosa (**Figure 4A** and **4E,F**): a luminal layer with reads
341 found beyond the cellular layer and likely representing cellular debris trapped in the mucus; three
342 apical layers expressing programs related to normal epithelial function (transmembrane transport,
343 oxidation-reduction process) with gradual transition from apical to basal features; and a basal-most
344 layer, enriched for the deep crypt, proliferation (G1/S,G2/M), MHCII and basolateral plasma
345 membrane programs, all common features of the deep crypt area. Finally, region 2, enriched with
346 fibroblasts, myofibroblasts and endothelial cells, and located in the most basal side of the tissue,
347 captured the submucosal and muscularis propria layers, which are predominantly comprised of
348 fibroblasts and muscle, respectively, alongside blood and lymphatic vessels, nerves and immune
349 cells. Overall, TACCO recovered the known organization of the colon, showing the power of our

350 unsupervised mapping approach and shedding light on expression programs that are required for
351 the maintenance of normal colon homeostasis.
352
353 **Dysplastic lesions maintain some of the programs of the corresponding regions in healthy**
354 **tissue**
355 AV lesions did not maintain the robust organization of normal tissues, and reflected the expected
356 histopathology of high grade dysplasia, when dysplastic cells are confined to the mucosal layer
357 and do not invade the submucosa[51] (**Figure 1A** and **4A**, top). Specifically, the submucosal and
358 muscularis propria layers from both normal and AV lesions were assigned to region 2 (**Figure**
359 **4A**).
360 Despite the altered morphology, some of the disrupted regions also expressed programs
361 characteristic of their normal healthy function, suggesting that tumor progression is spatially
362 structured and compartmentalized. For example, the region above the submucosa, captured as
363 region 1 in AV lesions and region 5 in normal colon (**Figure 4A**), had similar features in both AV
364 lesions and normal samples. Thus, although the overall spatial organization was disrupted in the
365 lesion, region 1 in AV lesions expressed programs that are reminiscent of the normal deep crypt
366 region 5, and was enriched for deep crypt cells and programs that are related to proliferation and
367 MHC II[43] (**Figure 4A,C**). These included proliferation programs 3 and 11 and both normal stem
368 cells (Epi03) and dysplastic secretory-like cells (Epi05), as well as dysplastic stem cells (Epi01,
369 though to a lesser extent than some other regions), so it may reflect one of the proliferative stem
370 cell (and dysplastic secretory-like) niches in AV models (**Figure 4B,C**). Other regions in the AV
371 lesions also contained some epithelial cells with normal profiles, expressing programs that should
372 allow them to maintain their capacity to perform normal tasks. For example, region 3 expressed

373 apical plasma membrane functions and region 10 was enriched with oxidation-reduction functions
374 (**Figure 4C**).
375 To learn about the spatial distribution of the dysplastic regions, we measured their distance from
376 region 2 (muscularis), which is a stable landmark in the lesions. Remnants of the layered structure
377 of the healthy tissue were still observed in the AV tissue, especially at relatively low distances
378 from the muscularis. For example, healthy region 5 – characteristically located at distances of
379 about 150-200µm from the muscularis – is replaced by dysplastic region 1, peaking at 200µm. All
380 malignant-like regions (6/8/11) were spatially associated at ~300-700µm from the muscularis
381 (**Figure S8C**), located ~100-400µm apart from each other (**Figure S8D**). We further validated this
382 result at the protein level, by staining for b-catenin, showing an (inactive) cytoplasmic localization
383 in the region adjacent to the muscularis, and mostly nuclear (active) localization in distal regions,
384 near the lumen (**Figure S8B**).
385
386 **Three spatially and functionally distinct tumor regions enriched in AV lesions**
387 Three regions – 6, 8, and 11 – had epithelial composition and programs that suggested advanced
388 malignant-like characteristics, each highlighting a potentially different mechanism for tumor
389 progression (**Figure 4G,H**). These three 'malignant-like regions', were enriched (*vs*. all other
390 regions) in stem cell, Wnt signaling and angiogenesis programs (16, #4 and #14; FDR=9.6 10^{-21},
391 8.1 10^{-10}, 3.7 10^{-10}, two-sided Welch's t-test on CLR transformed compositions) and depleted of
392 normal epithelial programs (#5, #8, and #10; FDR=5.5 10^{-12}, 1.6 10^{-12}, 4.2 10^{-10}, **Figure 4I**).
393 Furthermore, the malignant-like regions were enriched in immune cells, including monocytes-
394 macrophages (FDR<=1.5 10^{-5}; excluding Mac02 (Lyve1+)), T cell subsets TNK02 (Th1/Th17),
395 TNK05 (GdT/Il17+), TNK06 (Treg), TNK08 (proliferating T) (FDR=2.4 10^{-3}, 2.2 10^{-4}, 1.7 10^{-3},

396 9.6 10^{-4}; two-sided Welch's t-test on CLR transformed compositions), infiltrating granulocytes
397 (FDR<=9.7 10^{-3}), and mast cells (FDR=1.4 10^{-2}), suggesting an ongoing immune response
398 (**Figure 4J**). However, each one of the three regions had a different epithelial program
399 composition, suggesting that in each type of region there is a different dominant pathway/feature
400 that may drive tumor progression (**Figure 4C, Table S2**).
401
402 Region 6 was characterized by an inflammatory and angiogenic multicellular community, with
403 epithelial and immune cells expressing inflammatory programs, endothelial cells and monocytes
404 connected in a pro-angiogenic circuit, and pro-invasive genes expressed by both endothelial and
405 immune cells (**Figure 4B,C**). Specifically, region 6 was distinctly enriched for proliferation
406 (programs 3 and 11; FDR=2.2 10^{-14}, 1.2 10^{-15}, two-sided Welch's t-test on CLR transformed
407 compositions) and inflammatory epithelial programs (programs 6 and 7; FDR=9.0 10^{-7}, 2.0 10^{-}
408 ^{11}), and its non-epithelial compartment was correspondingly enriched for genes from
409 inflammatory pathways, including the response to TNF, IL-1 and IFN γ (FDR=3.1 10^{-4} , 2.8 10^{-3},
410 4.9 10^{-6}, two-sided Fisher's exact test in GO term enrichment), and chemotaxis of monocytes,
411 neutrophils, and lymphocytes (FDR=1.5 10^{-3}, 3.2 10^{-10}, 1.0 10^{-2}, **Table S2**), suggesting
412 recruitment of inflammatory cells from the circulation or other parts of the tissue. Region 6 was
413 also enriched for collagen binding genes and collagen-containing extracellular matrix (ECM)
414 genes (FDR=1.4 10^{-2}, 7.6 10^{-5}, two-sided Fisher's exact test in GO term enrichment, **Table S2**),
415 which are important for migration and invasiveness[52]. These include *Sparc*, expressed mainly by
416 endothelial cells and fibroblasts in our data, known to promote colorectal cancer invasion[53]; and
417 *Ctss*, a peptidase expressed by T cells and monocytes-macrophages that promotes CRC
418 neovascularization and tumor growth[54]. Finally, gene expression patterns in endothelial cells and

419 monocytes in region 6 suggested active angiogenesis through a multi-cellular feedback loop, with
420 enriched numbers of vascular and lymphatic endothelial cells expressing immune-attracting
421 chemokines (*Cxcl9*) and adhesion molecules (*e.g. Chd5, Mcam*), monocytes expressing
422 proangiogenic factors that induce proliferation of endothelial cells (*e.g.*, *Mmp12*), and monocytes
423 and macrophages expressing *Ctsd*, which increases tumorigenesis in CRC models[55] (**Figure S8E**).
424
425 Region 8 was enriched for deep crypt cells (program 13; FDR=1.7 10^{-14}, two-sided Welch's t-test
426 on CLR transformed compositions), reminiscent of the normal stem cell niche in normal colon, an
427 epithelial innate immune program (program 1; FDR=3.4 10^{-100}) expressed by secretory cells in AV
428 and AKPV lesions, and plasma and B cell activity. Unlike the canonical (normal) deep crypt region
429 (region 5), which is enriched for MHCII expression (program 18; FDR=8.6 10^{-28}), this region was
430 depleted for the program's expression (FDR=2.1 10^{-26}), which may indicate an earlier stem cell-
431 like state[43], or a decoupling of the cell cycle and MHCII programs (which are coupled in normal
432 ISC differentiation, and allow a cross talk with T cells to modulate T cell differentiation) (**Figure**
433 **4B,C**). The region's non-epithelial compartment was enriched for B cell activation and BCR
434 signaling genes (FDR=4.0 10^{-4}, 1.9 10^{-3}, two-sided Fisher's exact test in GO term enrichment,
435 **Table S2**). This may be related to B cell function in protection from lumen antigens[56] or to tertiary
436 lymphoid structures (TLS), which is correlated with clinical benefits in cancer patients[57]. Notably,
437 Epi05 (dysplastic secretory-like) enriched in Region 8 (**Figure 4B**) expressed higher levels of
438 inflammatory genes and immune chemokines (*e.g.*, *Ccl9, Ifitm3*) compared to normal counterparts,
439 Epi04 (secretory) (**Figure 2C** and **Figure S4C,E**), and may thus promote the formation of this
440 region. We validated the presence of TLS-like structures in association with deep crypt secretory
441 cells in AV lesions using multiplex RNA analysis (Cartana) (**Supplementary item 3**), showing

442 that the dominant population of B cells is accompanied by monocyte-macrophages and T cells
443 characteristic of TLSs, as well as the expression of *Reg4* and *Muc2*, deep crypt goblet/secretory
444 cell markers.
445
446 Region 11 was populated by cells expressing the Wnt signaling pathway program (4, FDR=9.3 10^{-}
447 ^{13}, two-sided Welch's t-test on CLR transformed compositions), with several lines of evidence
448 supporting an active epithelial to mesenchymal transition (EMT) in this region. Epithelial cells in
449 region 11 were enriched for the expression of mesenchymal genes, including Vimentin[58] (*Vim*,
450 FDR=7.4 10^{-245}, one-sided Fisher's exact test), *Prox1*[59] (FDR=3.7 10^{-153}), and *Sox11*[60] (FDR=7.7
451 10^{-224}) (**Figure S8F**), as well as for EMT signatures from a mouse model of lung adenocarcinoma
452 [61] (FDR=1.5 10^{-142}, two-sided Mann-Whitney U test) and from human head and neck squamous
453 cell carcinoma tumors[62] (FDR=6.9 10^{-55}, two-sided Mann-Whitney U test). This is consistent with
454 the role of Wnt signaling in promoting EMT and a mesenchymal phenotype in CRC, breast cancer
455 and other epithelial tumors[63,64]. Region 11 non-epithelial cells also expressed genes encoding
456 MHC-I binding proteins (FDR=4.6 10^{-2}, two-sided Fisher's exact test in GO term enrichment,
457 **Table S2**) and actin cytoskeleton, filament and binding proteins (FDR=3.1 10^{-6}, 7.7 10^{-3}, 2.7 10^{-}
458 ^{10}). Organization of the cytoskeleton affects migration, adherence, and interaction of lymphocytes
459 with antigen presenting cells[65]. Notably, region 11 also concentrated at a more distal part of the
460 tissue at ~900μm from the muscularis suggesting an outgrowth of the tissue towards the lumen
461 (**Figure S8C**).
462 Non-epithelial cells formed two cellular hubs in the malignant-like regions (6,8,11) (**Figure S8G**):
463 An endothelial-fibroblast hub, detected in all three regions, and an immune hub with B cells, TNK
464 cells, monocytes, and macrophages, which was prominent in inflammatory region 6, weaker (less
465 spatially correlated) in region 8 (but validated *in situ*), and not correlated in region 11. Thus,

466 activation of an immune response is reflected by close proximity between immune cells. We
467 further characterized the organization of the vascular niche using our multiplex *in situ* RNA data,
468 finding that while neighbors of the Pdgfrb-expressing pericytes are mainly other Pdgfrb-
469 expressing pericytes and endothelial cells, Pecam1-expressing endothelial cells appear self-
470 enriched next to themselves at cellular scale distances and close to Pdgfrb-expressing pericytes
471 for larger distances (**Figure S8H**).
472
473 Overall, three multicellular community regions were enriched in AV lesions: (**1**) inflammatory
474 epithelial regions with endothelial cells and monocytes expressing angiogenesis, inflammation and
475 invasion programs; (**2**) epithelial stem-like regions, associated with plasma and B cell activity; and
476 (**3**) regions with epithelial to mesenchymal transition (EMT) and Wnt signaling dysplastic cells.
477 Each region highlights different processes that modulate tumorigenesis or invasion, and the three
478 regions co-exist in the same tumor at different spatial locations.
479
480 **Cell-cell interactions are rewired in AV lesions**
481 To identify cell-cell signaling mechanisms that may underly these regional associations, we used
482 COMMOT[66], a computational framework that uses Optimal Transport to infer cell-cell
483 communication from receptor-ligand expression patterns in spatially resolved data. We used
484 COMMOT's bead-wise communication 'output' and devised a method to address p-value inflation
485 in statistical enrichment testing, using spatially-informed data aggregation (**Methods**).
486
487 We observed stronger and distinct ligand-receptor interactions in AV *vs*. normal samples,
488 reflecting the activated state in the dysplastic tissue (**Figure 4K, Supplementary item 4A**). In

489 particular, while interactions enriched in AV lesions involved immune, epithelial and stromal
490 signaling, those enriched in normal tissue involved neuropeptides, such as NPY and GCG
491 (glucagon). Moreover, malignant-like regions 6 and 11 as well as region 2 (muscularis) were
492 particularly enriched for active communication pathways (**Figure S8I** and **Supplementary item**
493 **4B**). This included the WNT signaling pathway, angiogenesis (VEGF, PDGF, FGF) and the OSM
494 pathway.
495

496 **Similar spatial patterns in human and mouse tumors**

497 The overall spatial distribution of cell types and epithelial profiles was conserved between mouse
498 and human tumors, when comparing to scRNA-seq[2,28,29,67–70]. We examined mouse-defined
499 regions in human tumors, using TACCO to map the expression profiles associated with the
500 epithelial, immune and stroma compartments in each of the TACCO-identified mouse regions to
501 scRNA-seq profiles from human CRC, and probabilistically annotated region-specific expression
502 profiles for each scRNA-seq profile from the human samples. This identified two main "meta
503 compartments", with epithelial, stromal and immune profiles from human MMRp and MMRd
504 tumors associated with regions 6, 8 and 11 that were enriched in AV lesions (as well as 0 and 2),
505 while those from normal human tissue were associated with normal regions (*e.g.*, 5, 10, 12)
506 (**Figure 5A** and **Figure S9A**, **Methods**).
507

508 **Malignant-like regions are associated with tumor progression in human colorectal tumors**

509 We next assessed if the regional epithelial programs that we spatially identified in mouse are
510 conserved in human. To this end, we constructed pseudo-bulk profiles from epithelial cells for our
511 mouse samples and for recently published human samples profiled along different stages of

512 malignant transformation, from normal tissue to polyp to CRC[2,28,29,67–70]. We scored each epithelial
513 pseudo-bulk profile with the differentially expressed genes between the epithelial parts of the
514 regions and computed the principal components of these scores across all human and mouse
515 samples (**Methods**). The first principal component (PC1) captured features that are related to
516 malignancy, with higher values for human tumors *vs*. polyps (**Figure 5B** and **Figure S9B**). In
517 addition, malignant-like region (6/8/11) scores were higher in dysplastic *vs*. normal samples
518 (**Figure 5C**). Thus, the spatial region profiles defined in mouse capture features that correlate with
519 malignant transformation in human.
520
521 We further classified each full pseudo-bulk profile from the dysplastic human samples into one of
522 the four groups in the CMS expression-based classification[71] (**Methods**) and compared the mouse
523 region scores for each class of samples (**Figure S9C, Methods**). CMS2 classified samples were
524 most closely related to our dysplastic mouse models: all of the dysplasia -associated regions were
525 enriched in CMS2 tumors while most normal regions were depleted, relative to the other CMS
526 classes.
527
528 Finally, we found that expression of the malignant like regions (6/8/11) in tumors was associated
529 with clinical outcome. We scored each tumor based on genes that were differentially expressed
530 between the full expression profile of malignant-like regions (6/8/11) and compared the
531 progression-free interval (PFI) and overall survival (OS) for patients in TCGA whose RNA-seq
532 profiles were in the top and bottom quartile of malignant-like region scores (**Methods**). High
533 scores for malignant-like region 11 (EMT) were correlated with shorter PFI, while those for
534 malignant-like region 6 (inflammation) correlate with longer PFI and longer OS, (**Figure 5D,E**).

535 These associations were driven primarily by MMRp tumors **(Figure S9D,E**). This suggests that
536 region 11 is associated with pro-tumorigenic properties in human patients, while region 6 might
537 be associated with tumor controlling properties. This highlights the importance of multicellular
538 functional tissue modules in the CRC tumor ecosystem.

539

540 **DISCUSSION**

541

542 Here, we systematically charted the spatial organization of cellular expression in dysplastic tissue
543 of the colon, to help identify putative functional units in the tumor. We used TACCO[21] to integrate
544 scRNA-seq and Slide-seq data, not only by mapping cell types to their positions, but also
545 distinguishing different cell programs, the regions that they dominate, and their characteristic
546 microenvironments. This allowed us to overcome technical limitations, such as lack of spatial
547 context in scRNA-seq and sparse readout in Slide-seq, and to generate a high-resolution spatial
548 map of the dysplastic landscape transcending beyond the mapping of individual cells to spatial
549 positions. We used this map to show correlation with clinical outcome in human patient tumors.

550

551 Our scRNA-seq analysis revealed profound enrichment of a stem cell program in dysplastic
552 tissues. The profiles of dysplastic cells expressing this program are distinct from normal stem cells
553 and enriched with expression of negative regulators of the WNT signaling pathway and
554 inflammation, suggesting a non-canonical function. The abundance of these cells with stemness
555 potential across all our malignant-like regions, points to a dynamic population that can affect the
556 cells in its proximity, by secretion of negative regulators of the WNT signaling and inflammatory
557 function but may also adopt various functions depending on the environmental cues and dysplasia-

558 associated cells in its proximity. A similar population, designated “high-plasticity cell state”, was
559 previously described in a mouse model of lung adenocarcinoma and in human patients, where it
560 was correlated with resistance to chemotherapy[61]. Whether these cells can be manipulated to take
561 on specific phenotypes or even to differentiate into normal-like enterocytes given the appropriate
562 signal from the microenvironment, remain as open questions.
563
564 Within the dysplastic lesions, alongside malignant-like regions, we found regions with normal
565 features (Regions 3,4,9,10), comparable to regions found in the normal colon, most likely
566 representing compartments driven by clones that were not affected by the genetic perturbation.
567 One of these regions, region 4, contained mainly goblet cells with normal expression profiles.
568 Whether this neighborhood represents normal cells that reside alongside malignant cells or a
569 cancer transition state, it may modify tumor progression, by recruiting immune cells or by
570 secreting factors that affect epithelial proliferation in adjacent regions. For example, region 4 in
571 dysplastic lesions is enriched with chemokine activity genes relative to region 4 in normal colon
572 suggesting a possible role in recruitment of immune cells to the dysplastic landscape. Further work
573 is required to understand the role of these regions (expressing normal features) in tumor
574 progression.
575
576 While the malignant-like regions were identified as discrete spatial entities, each with coordinated
577 features across epithelial, immune, and stromal cells, these regions are adjacent to each other.
578 Thus, they may still influence one another through signaling or by utilizing branches of the same
579 main vessels. For example, *Osm* is expressed by cells in region 6, whereas its receptor is expressed
580 on fibroblasts and endothelial cells enriched in region 11. *OSMR* was previously shown to be

581 expressed by inflammatory fibroblasts[44,72] and its activation in malignant cells promotes EMT in
582 breast cancer and pancreatic cancer[73,74] and a mesenchymal state in glioblastoma[75]. Future studies
583 can help determine if these regions are functionally inter-dependent and if they evolved from the
584 same clones and can inter-convert, or whether they developed independently.
585
586 Because animal models complement cell and tissue atlases of human colorectal cancer[2,28], by
587 allowing experimental manipulation for mechanistic studies[76,77], it is important to relate between
588 models and patients. By studying genetically engineered mouse models using high resolution
589 single cell spatial genomics we can help determine to what extent they recapitulate the cellular and
590 spatial organization of human disease, in the context of two distinct genetic states that represent
591 human CMS-2 lesions. To this end, we developed several approaches to allow cross species
592 comparison of tumors at the single cell and spatial level, despite the high level of both intra- and
593 inter-individual variation within each species. Comparing to human CRC, our analysis suggests
594 that the CRC landscape is organized in similar multicellular functional tissue modules between
595 human and mouse, and disease subtypes (*e.g.*, MMRp and MMRd). Future studies applying our
596 approaches to patient cohorts could help understand whether the expression of different tissue
597 modules may contribute to the partial response to immunotherapy reported for MMRd patients[78],
598 and to define specific tissue modules predictive of response to therapy. Notably, while our study
599 focused on the tumor landscape, its findings may be relevant for tissue response to other challenges
600 (*e.g.*, inflammation, fibrosis, wound healing), which involve activation of similar functional tissue
601 modules, a result of collective function of parenchymal, immune and stromal cells.

602 Taken together, our integrative approach facilitates spatial analysis with high resolution,
603 constructing regional neighborhoods and their spatial layout at both high cellular resolution and

604 genomic scale. Our work is an important step toward a systematic understanding of the
605 organization of dysplastic tissue with the potential to contribute to improved patient stratification
606 by the multicellular functional units in the tumor landscape.

607

608 **Methods**

609 **Human subjects**

610 The MGH Institutional Review Board approved protocols for tissue collection used for sequencing
611 (Protocol 02-240). Informed consent was obtained from all subjects prior to collection. Only
612 patients with primary treatment- naive colorectal cancer were included in this study.

613

614 **Mice**

615 Mice were housed in the animal facility at the Koch Institute for Integrative Cancer Research at
616 MIT. All animal studies described in this study were approved by the MIT Institutional Animal
617 Care and Use Committee (Protocol 1213-106-16). *Apc*$^{fl/fl}$ mice[79] were obtained from NCI mouse
618 repository, *Kras*$^{LSL\text{-}G12D/+}$ Ref.[80], *Rosa26*$^{LSL\text{-}tdTomato}$ Ref.[81] and Trp53$^{fl/fl}$ Ref.[82] mice obtained from
619 Jackson, *Villin*^{CreERT2} Ref.[83] mice were a gift from Dr. Sylvie Robine. All mice were maintained
620 on C57BL/6J genetic background. Approximately equal numbers of male and female mice of 6–
621 10 weeks of age were used for all experiments. Where indicated, mice were injected to the
622 submucosal layer of the colon with 4-hydroxytamoxifen (EMD Millipore # 579002) dissolved in
623 ethanol at a concentration of 100 μM (for the mice that were kept for 3 weeks after injection) or
624 30 μM (for the mice that were kept for 9 weeks after injection). Tumors were resected at either 3
625 or 9 weeks after 4-hydroxytamoxifen injection. Colonoscopy and colonoscopy-guided injection
626 methods were previously described in detail[30,31].

627

628 **Tissue processing for scRNA-seq**

629 Single-cell suspensions from healthy colon or dysplastic lesions were processed using a modified
630 version of a previously published protocol[44]. Tissue samples were rinsed in 30ml of ice-cold PBS

631 (ThermoFisher 10010-049), chopped to small pieces and washed twice in 25 ml PBS, 5mM EDTA
632 (ThermoFisher AM9261), 1%FBS (ThermoFisher 10082-147). To prime tissue for enzymatic
633 digestion, samples were incubated for 10 minutes at 37°C, placed on ice for 10 minutes before
634 shaking vigorously 15 times followed by supernatant removal. Tissues were placed into a large
635 volume of ice-cold PBS to rinse prior to transferring to 5ml of enzymatic digestion mix (Base:
636 RPMI1640, 10 mM HEPES (ThermoFisher 15630-080), 2% FBS), freshly supplemented
637 immediately before use with 100 mg/mL of Liberase TM (Roche 5401127001) and 50 mg/mL of
638 DNase I (Roche 10104159001), and incubated at 37°C with 120 rpm rotation for 30 minutes. After
639 30 minutes, enzymatic dissociation was quenched by addition of 1ml of 100% FBS and 10mM
640 EDTA. Samples were then filtered through a 40 mM cell strainer into a new 50 mL conical tube
641 and rinsed with PBS to 30 mL total volume. Tubes were spun down at 400 g for 7 minutes, at 4°C.
642 Resulting cell pellets were resuspended in 1ml PBS, placed on ice and counted.
643
644 **Cell hashing**
645 Cell hashing was performed based on the published protocol[84] as summarized below. Dissociated
646 cells were resuspended in 1ml of Cell Hashing Staining Buffer (1× PBS with 2% BSA (New
647 England Biolabs, B9000S) and 0.02% Tween (Tween®-20 Solution, 10%, Teknova, VWR-
648 100216-360) and counted. 500,000 cells were resuspended in 100 µL of Cell Hashing Staining
649 Buffer and incubated for 30 minutes on ice, with 2 µL of the appropriate BioLegend TotalSeq™
650 Hashing antibody (a 1:50 dilution, using a total of 1 µg of antibody per cell suspension).
651 TotalSeq™-A anti-mouse Hashtag antibodies #1-8 (catalog numbers:155801, 155803, 155805,
652 155807, 155809, 155811, 155813, 155815) were used. Cells were washed three times with 0.5 mL
653 of Cell Hashing Staining Buffer and filtered through low-volume 40-µm cell strainers. All cell

654 suspensions were recounted to achieve a uniform concentration of 7,000 cells per microliter before
655 pooling for capture by 10x Chromium controller following the manufacturer protocol for the v2 or
656 v3 3' kit (10X Genomics, Pleasanton, CA).
657
658 **Hashtag oligo (HTO) library preparation**

659 Separation of hashtag oligo (HTO)-derived cDNAs (<180 bp) and mRNA-derived cDNAs
660 (>300 bp) was done after whole-transcriptome amplification by performing 0.6× SPRI bead
661 purification (Agencourt) on cDNA reactions as described in 10x Genomics protocol. Briefly,
662 supernatant from 0.6× SPRI purification contains the HTO fraction, which was subsequently
663 purified using 1.4 and 2× SPRI purifications per the manufacturer's protocol (Agencourt). HTOs
664 were eluted by resuspending SPRI beads in 15 µL TE. Purified HTO sequencing libraries were
665 then amplified by PCR (1µL clean HTO cDNA, 25µL 2X NEBNext Master Mix (NEB #M0541)),
666 10 µM SI-PCR and D701or D704 primers performed dial out PCR (98°C (10 sec), (98°C for 2 sec,
667 72°C for 15 sec) x 12/18 then 72°C for 1 min) for 12 and 18 cycles, and used the 18 cycles product
668 for sequencing. PCR reactions were purified using another 2× SPRI clean up and eluted in 15 µL
669 of 1× TE. HTO libraries were quantified by Qubit High sensitivity DNA assay (ThermoFisher)
670 and loaded onto a BioAnalyzer high sensitivity DNA chip (Agilent).

671 SI-PCR: AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGC*T*C
672 D701:
673 CAAGCAGAAGACGGCATACGAGATCGAGTAATGTGACTGGAGTTCAGACGTGTGC
674 D704 :
675 CAAGCAGAAGACGGCATACGAGATGGAATCTCGTGACTGGAGTTCAGACGTGTGC
676

677 **Sequencing**

678 Samples were sequenced using HiSeq X (Illumina). Hashing libraries were sequenced with spike-
679 ins of 2.5%.

680

681 **Tissue processing for Slide-seq**

682 Colons were flashed with cold PBS and a segment that includes the lesion and surrounding tissue
683 (or a respective healthy segment from normal mice) was dissected. Samples were then mounted in
684 cold OCT, flash frozen on dry ice covered with ETOH and long-term stored in -80°C.

685

686 **Slide-seq**

687 For mouse and human experiments, 10 μm sections were cut and the Slide-seq V2 protocol was
688 used as previously described[15]. For mouse experiments, four and six arrays were collected from
689 normal colons and AV lesions respectively. The muscularis was fit onto the array of both healthy
690 and dysplastic lesions to allow appropriate orientation.

691

692 **Multiplex *in situ* RNA Analysis**

693 Multiplex *in situ* RNA analysis was performed with Cartana[33] technology (a newer version is now
694 available as Xenium (10x Genomics)). In total we measured 3 samples with one section per sample
695 in each state (normal/AV) and probe set (V1/V2), with an additional replicate section for one of
696 the samples in normal V2.

697

698 Fresh Frozen OCT-embedded tissues from normal colon and AV lesions were cryosectioned as 10
699 μm sections and placed onto SuperFrost Plus glass slides (ThermoFisher) and further stored at -

700 80˚C before experiments. Samples were fixed (with 4% formaldehyde) for 5 minutes and
701 permeabilized for 5 minutes (with 0,1 mg/mL pepsin in 0.1M HCl (P7012 Sigma-Aldrich)) prior
702 to library preparation.
703
704 For library preparation, chimeric padlock probes (directly targeting RNA and containing an anchor
705 sequence as well as a gene-specific barcode) for a custom panel of 66 (V1) or 180 (V2) genes
706 (**Table S3**, see below) were hybridized overnight at 37°C, then ligated before the rolling circle
707 amplification was performed overnight at 30°C using the HS Library Preparation kit for
708 CARTANA technology and following manufacturer's instructions. All incubations were
709 performed in SecureSeal^{TM} chambers (Grace Biolabs). Note that prior to final library preparation,
710 optimal RNA integrity and assay conditions were assessed using Malat1 and Rplp0 housekeeping
711 genes only using the same protocol.
712
713 To prevent tissue sections detachment, an additional baking step of 30 minutes at 37°C was
714 performed before mounting. To quench autofluorescence background, TrueView (SP-8400
715 VectorLabs) was used for 1 minute at room temperature. For tissue sections mounting, Slow Fade
716 Antifade Mountant (Thermo Fisher) was used for optimal handling and imaging.
717
718 Quality control of library preparation was performed by applying anchor probes to detect
719 simultaneously all rolling circle amplification products from all genes in all panels. Anchor probes
720 are labeled probes with Cy5 fluorophore (excitation at 650 nm and emission at 670 nm).
721 All samples passed quality control and went through *in situ* barcode sequencing, imaging and data
722 processing. Briefly, adapter probes and sequencing pools (containing 4 different fluorescent labels:

723 Alexa Fluor® 488, Cy3, Cy5 and Alexa Fluor® 750) were hybridized to the padlock probes to
724 detect the gene-specific barcodes, through a sequence specific signal for each gene specific rolling
725 circle amplification product. This was followed by imaging and performed 6 times in a row to
726 allow for the decoding of all genes in the panel. To reduce lipofuscin autofluorescence, 1X
727 Lipofuscin Autofluorescence Quencher (Promocell) was applied for 30 seconds before
728 fluorescence labeling.
729
730 Raw data consisting of 20x or 40x images from 5 fluorescent channels (DAPI, Alexa Fluor® 488,
731 Cy3, Cy5 and Alexa Fluor® 750) were each taken as z-stack and flattened to 2D using maximum
732 intensity projection. After image processing and decoding, the results were summarized in a csv
733 file and gene plots were generated using MATLAB [85].
734
735 **scRNA-seq pre-processing and quality control filtering**
736 Count matrices for scRNA-seq were generated using the Cumulus feature barcoding workflow
737 v0.2.0[86] with CellRanger v3.1.0 and the mm10_v3.0.0 mouse genome reference. Cell profiles were
738 quality filtered by requiring between 1,000 and 50,000 counts, and between 500 and 7,000 genes,
739 less than 20% mitochondrial counts, and less than 10% hemoglobin counts. Cell profiles that did
740 not meet all these criteria were discarded. The top 5,000 highly variable genes were annotated on
741 the remaining cells after normalization to 10,000 counts and log1p transform using Scanpy's
742 "highly_variable_genes" function[87] and providing the chemistry (v2/v3) by hashing (True/False)
743 combination as batch-annotation. Putative doublets were removed using Scrublet[88] with default
744 parameters.
745

746 **Selection of variable genes, dimensionality reduction and clustering**

747 A preliminary clustering using the Leiden algorithm with resolution 1.0 was performed after
748 normalization to 10,000 counts, log1p transform, correction for number of counts and percentage
749 of mitochondrial genes, scaling with a max_value of 10, and generating a *k*-nearest neighbors (*k*-
750 NN) graph with 15 neighbors on a PCA of the previously annotated 5000 highly variable genes
751 with 50 components using Scanpy[87]. The single cell profiles were provisionally annotated with
752 SingleR[89] cell-wise (i.e. without using clustering information) using the SingleR built-in
753 MouseRNAseqData and an intestine specific dataset from Tabula Muris[90]
754 (https://figshare.com/ndownloader/files/13092143). For further processing, the dataset was then
755 split into the three compartments, epithelial, immune and stromal, using the provisional SingleR
756 annotations.

757 For each compartment, the top 5,000 highly variable genes were annotated using Scanpy's
758 "highly_variable_genes" function on cells normalized to 10,000 counts after log1p-transformation
759 and providing the chemistry (v2/v3) by hashing (True/False) combination as batch-annotation.

760

761 **Expression programs and batch correction**

762 For the dataset of each compartment separately (generated as described above), an integrative
763 NMF was performed (using a part of the LIGER[45] implementation) with k=20 and lambda=5 to
764 identify 20 programs and their respective weights per cell. This iNMF factorization represents the
765 single cell expression matrix as a weighted sum of profiles such that both the weights and programs
766 contain only non-negative numbers, while allowing for and separating out batch-only
767 contributions. The same approach was also used with a higher *k* (epithelial and immune: 200,
768 stromal: 50) to yield a detailed and batch corrected decomposition of expression which was then

769 combined to obtain a count-like corrected expression matrix for the top 5,000 highly variable
770 genes. For each compartment separately, these batch-corrected data were normalized to 10,000
771 counts, log1p transformed, corrected for number of counts and percentage of mitochondrial genes
772 by linear regression, scaled with a max_value of 10, followed by a PCA of the previously annotated
773 5,000 highly variable genes. A *k*-nearest neighbors (*k*-NN) graph was constructed from the top 50
774 PCs, with *k*=15 neighbors using Scanpy, and clustered using a compartment-specific Leiden
775 resolution parameter (epithelial: 0.2, immune: 0.4, stromal: 0.1). This clustering was used as the
776 cluster level annotation of the mouse scRNA-seq data for the epithelial and stromal compartment.
777 Separately per compartment the data were annotated with SingleR using the cluster information.
778 The same per-compartment batch-corrected and preprocessed data from the Leiden clustering was
779 used to create UMAP embeddings with PAGA initialization using Scanpy.

780

781 To improve the clustering and annotation in the immune compartment and to filter out additional
782 doublets not detected by Scrublet, the immune data were separately filtered and clustered using
783 information from the compartment level clustering and annotation. To that end, myeloid and T/NK
784 cells were partitioned separately and further processed, and additional likely doublet cells were
785 labeled and removed by the following procedure:

786 1. Cells were labeled as doublets based on higher number of UMIs of marker genes for other
787 compartments than the 95th percentile observed in this immune partition (*i.e.*, Epcam and Cdh1
788 to remove immune-epithelial doublets and Cav1 and Kdr to remove immune-stromal doublets)
789 and other immune partitions (i.e., Cd3d, Cd3e, and Cd3g to remove myeloid-lymphoid
790 doublets from the myeloid cells). This type of filter criterion for lowly expressed genes ("larger

791 than some percentile" on integer counts) also allows to keep more than 95% of the cells if, for
792 example, all cells of this partition happened to have 0 UMIs of a particular marker gene.
793 2. Cells were labeled as doublets if they had inconsistent cell-wise and cluster-wise SingleR
794 annotations.
795 3. Cells were labeled as doublets if they had significantly (Benjamini-Hochberg FDR=0.05, one-
796 sided Fisher's exact test) more neighbors in the k-NN graph from the immune compartment
797 that were already marked as doublets.
798 4. All cells labeled as doublets were removed.

799 After filtering, the count matrices were batch corrected as above using the integrative NMF from
800 LIGER with k=20 and lambda=5, and clustered like above with group specific Leiden resolution
801 (myeloid: 0.2, TNK: 0.4). For myeloid and TNK cells, this clustering superseded the original
802 clustering. The integrative NMF result here was only used for updating the clustering and not for
803 generating an extra set of expression programs.

804 Note that cluster Epi06 shows a broad expression spectrum; while such pattern can best be
805 explained by remaining doublets, they were not called by Scrublet and could also not be
806 consistently removed by extra QC as was applied to the immune compartment. The interpretation
807 of Epi06 as containing doublets is supported by its overrepresentation in the annotation of the
808 Slide-seq data (Figure S6E), as both doublets and slide-seq beads represent compositions of
809 different cells. Still, non-doublet explanations (e.g. undifferentiated cells) cannot be ruled out.
810 Therefore, Epi06 cells are retained in the data but not interpreted biologically.

811

812 **Marker selection for *in situ* RNA analysis (CARTANA)**

813 CARTANA V1 markers were selected from genes differentially expressed between compartments,
814 cell types and clusters (for the immune and stromal compartments), and highly ranked genes for
815 programs (for the epithelial compartment) were first filtered by biological relevance and the
816 literature, obtaining 87 genes. To further reduce the set to the available panel size (66 genes), genes
817 were annotated into categories of potential redundancy. To choose between redundant genes, a
818 global objective function was optimized over the gene selection, looping over all potentially
819 redundant gene sets until convergence, exhaustively testing all choices within a gene set, and
820 accepting the best choice for this gene set in terms of the global objective function. The global
821 objective function was constructed as the mean 4-fold cross-validation scores (using
822 "GroupKFold" and "cross_val_score" from sklearn[91]) of multi-class logistic regression
823 classification (using "LogisticRegression" from sklearn) for discriminating cell classes and of
824 ridge regression of epithelial program weights (using "RidgeRegression" from sklearn) on the
825 scRNA-seq data subsampled to the expected sparsity of CARTANA data. The cell classifications
826 used were between the stromal, immune, and epithelial compartment, within the stromal
827 compartment between Endo01, Endo02, and fibroblasts, within the immune compartment between
828 myeloid and lymphoid lineage, within the lymphoid lineage between T cells and B cells, within
829 the myeloid lineage between granulocytes, mast cells, and all monocytes and macrophages
830 together, between monocytes and macrophages, and within monocytes between Mono01, Mono02,
831 Mono03, and Mono04. The programs used in ridge regression were programs 3, 4, 6, 7, 13, 14,
832 15, and 16. As the probes for Ly6c1 and Ly6c2 could not discriminate sufficiently between Ly6c1
833 and Ly6c2, we chose combined probes that measure both.

834 The CARTANA V2 panel, included 59 of the 66 genes in the V1 panel (the others had to be
835 removed for technical reasons), another 113 genes from the standard fixed gene panel for
836 CARTANA, and 8 selected genes from literature.
837

838 **Analysis of CARTANA data**

839 Each measured molecule was annotated with an originating cell type cluster label (using TACCO's
840 "tc.tl.annotate_single_molecules", with RCTD[19] as the core annotation method and parameters
841 bin_size=20, n_shifts=3, assume_valid_counts=True) separately for each sample. For this, genes
842 in the reference that would likely cross hybridize in the probe panel design were summed over
843 (Ly6c1 and Ly6c2). TLS-like regions were annotated by visual inspection of the cell type cluster
844 composition and morphology.

845 To assess cell type compositions of the full dataset, molecules with cell type cluster annotations
846 were binned into 10µm bins (using TACCO's "tc.utils.bin" and "tc.utils.hash" functions) and
847 cluster-level annotations were merged to cell-type level.

848 To assess the compositions of TLS-like regions, CARTANA v2 data were aggregated, conserving
849 the categorical TLS annotation (using TACCO's "tc.utils.bin" and "tc.tl.dataframe2anndata"
850 functions).

851 **Comparison between experimental methods**

852 To compare cell type composition between methods, CLR-transformed compositions of samples
853 (or of spatially split samples for the spatial methods, see subsection "Ligand-Receptor Analysis
854 and Spatially Informed Enrichment") were computed. Then, using a 100 bootstrapped means of

855 the compositions, mean and standard error of the mean were calculated once for the AV and once
856 for the normal samples. To compare gene expression, the mean difference of gene counts between
857 normal and AV samples and its standard error were calculated using all pairwise differences
858 between the bootstrapped normal and AV samples. For gene comparisons, the CLR-transformed
859 composition over all genes that were measured in all of scRNA-seq, Slide-seq, CARTANA V1,
860 and CARTANA V2 was used. Pearson correlation between mean compositions was calculated for
861 each pair of methods.

862

863 **RNA-velocity analysis**

864 Splicing aware count matrices for scRNA-seq were generated using CellRanger v6.1.2 and
865 velocyto v0.17.17[37] with the ensembl v108 mouse genome reference. Scvelo v0.2.5 was used to
866 infer velocity separately for the epithelial and TNK subsets (using the functions
867 "scv.pp.filter_and_normalize", "scv.pp.moments", "scv.tl.velocity" (with mode='stochastic'), and
868 "scv.tl.velocity_graph"). Scanpy and bbknn v1.5.1[92] were used to generate batch-corrected UMAP
869 embeddings for the two subsets for visualization with scvelo's
870 "scv.pl.velocity_embedding_stream" function.

871

872 **Selection of human single cell data for the comparison of cell type and epithelial program**
873 **composition**

874 ScRNA-seq data from Ref[2] was used as reference for human CRC. To avoid biases in cell type
875 compositions, only the subset of the data where "PROCESSING_TYPE = unsorted" were used.

876

877 **Comparison of human and mouse samples by cell type composition**

878 To compare human and mouse samples by composition of T/NK cell subsets, T/NK annotations
879 from mouse and human data[2] were matched by TACCO, using optimal transport (OT). First,
880 human expression data were mapped to mouse genes using MGI homology information
881 [subsection "Mapping of mouse and human orthologs"]. Then, human cell cluster annotations
882 ('cl295v11SubFull') were mapped from the subset of human cells annotated as T/NK/ILC to the
883 subset of mouse cells annotated as T/NK using TACCOs "annotate" function with OT as core
884 method, basic platform normalization, entropy regularization parameter epsilon 0.005, marginal
885 relaxation parameter lambda of 0.1, and 4 iterations of bisectioning with a divisor of 3. Annotation
886 with maximum probability per cell was used as the unique cluster level annotation for mouse T/NK
887 cells. Annotations were aggregated per sample to yield a compositional annotation over the
888 identical cluster annotation categories (from the human dataset) for the T/NK subsets of human
889 and mouse samples. Annotation vectors were then processed using the sc.pp.neighbors and
890 sc.tl.umap functions from Scanpy[87] to yield a 2D sample embedding with respect to T/NK cell
891 composition. Using the coordinates in the UMAP in place of spatial coordinates, neighborship
892 enrichment z-scores were computed with TACCO's co_occurrence_matrix function with
893 max_distance=2 and n_permutation=100.
894

895 **Slide-seq compositional annotation**

896 Slide-seq data were annotated with scRNA-seq reference annotations. First, Slide-seq and scRNA-
897 seq data were filtered to retain only 15759 genes that were detected in both datasets and only beads
898 and cells that had at least 10 counts across all these common genes.
899

900 Next, TACCO[21] was used to perform compositional annotation of each bead, allowing the bead
901 counts to be explained by fractional contributions. In its basic application, TACCO finds an
902 "optimal" mapping between scRNA-seq annotation categories (*e.g.*, cell types) and beads by
903 solving a variant of an entropically regularized Optimal Transport (OT) problem in expression
904 space. In its iterative application, TACCO uses a bi-sectioning functionality iteratively, annotating
905 only fractions of the counts in each round, and reserving the remainder for the next round to
906 improve the sensitivity to sub-leading annotation contributions (that is, first capture a portion of
907 the counts for the "top" cell types, but preserving others for other, more minor, cell types).
908
909 For the compositional annotation of Slide-seq beads with the categorical cell clusters from the
910 single cell data, the "annotate" function of TACCO with OT was used as core annotation method
911 per Slide-seq puck with the subset of the single cell data with matching disease state, with basic
912 platform normalization, entropy regularization parameter epsilon 0.005, marginal relaxation
913 parameter lambda of 0.001, and 4 iterations of bisectioning with a divisor of 3.
914
915 For the compositional annotation of Slide-seq beads with the compositional epithelial programs,
916 the annotated beads were split using the "split_observations" function of TACCO on the cluster-
917 level annotation and then aggregated to compartment level using the "merge_observations"
918 function keeping only beads for a compartment with at least 50 counts assigned to that
919 compartment. The genes from the epithelial part were filtered to retain only those that were used
920 to define the epithelial programs in scRNA-seq, and then annotated using again the "annotate"
921 function with OT as core annotation method, basic platform normalization, entropy regularization
922 parameter epsilon 0.01, and a marginal relaxation parameter lambda of 0.001.

923

924 **Slide-seq region annotation**

925 Region annotation was done for all pucks (normal and AV pucks) in one step to get comparable
926 region annotations across pucks, such that beads that have similar epithelial program activity and
927 are spatially close are called as one region. Because cell type composition can change drastically
928 from one bead to its neighbor at the length scales of Slide-seq data, there is a need to compromise
929 between optimizing the two similarities. This is done with the "find_regions" function of TACCO,
930 which performs a Leiden clustering on the weighted sum of connectivity matrices derived from
931 epithelial program similarity and spatial proximity, using a position weight of 0.7, a Leiden
932 resolution of 1.3, and 15 nearest neighbors per bead in position space and epithelial program space.
933 To determine the neighbors in epithelial program space, the square-roots of the program weights
934 were used for neighbor finding which effectively uses Bhattacharyya coefficients as overlap in
935 epithelial program space instead of the Euclidean scalar products used for position space. These
936 regions are defined by construction only on beads with a large enough epithelial contribution (see
937 above) and are then extended to all beads by assigning unannotated beads the region from the
938 nearest bead with region annotation.

939 Submucosal and muscularis propria layers are predominantly comprised of fibroblasts and muscle
940 cells, respectively, alongside blood and lymphatic vessels, nerves and immune cells. Our algorithm
941 depends on the epithelial expression component in beads. Since these layers do not contain
942 epithelial cells[93], mapping likely relied either on "noisy" signal from non-epithelial cells or from
943 the basal-most epithelial layer.

944 To determine region composition at a certain distance of a reference region, TACCO's
945 "annotation_coordinate" function is used with max_distance=1000 and delta_distance=10.

946

947 **Slide-seq quality filtering**

948 For all downstream analyses, all beads with less than 100 reads were discarded.

949 **Region- and cell type- characterizing genes in Slide-seq data**

950 Genes to characterize regions on Slide-seq pucks irrespective of compartment composition were
951 found using Scanpy's rank_genes_groups function on the full bead expression profiles. To find
952 them separately for each compartment, the compartment-level split beads [sub-section "Slide-seq
953 annotation"] were used instead of the full beads. To compare gene expression between cell types
954 on Slide-seq pucks, cluster-level split beads [sub-section "Slide-seq annotation"] were aggregated
955 to cell type level.

956 **EMT scoring**

957 Malignant regions were scored for EMT signatures, using only counts attributed to the epithelial
958 compartments within these regions and only genes expressed on at least 3 beads. Bead profiles
959 were normalized to 10,000 counts, log1p transformed and scaled, and Scanpy's "sc.tl.score_genes"
960 function was used to score the top 50 genes in two EMT gene signatures[61,94].

961 **Cell-type neighborships in Slide-seq data**

962 To evaluate the local cell-type neighborship relations in the different disease states on the cluster
963 level, the clusters were filtered per disease state to contain only clusters which account for at least
964 1% of the UMIs in that state. Then neigbourhood-enrichment z-scores were calculated using
965 TACCO's "co_occurrence_matrix" function with max_distance=20 and n_permutation=10. To

evaluate the stability of the result, this is also repeated for (max_distance, n_permutation)=(40,10),(60,10),(20,5), and (20,50). To get the significance of the overall change in z-scores between the states, a one-sided Mann-Whitney-U test was performed on the values of the upper triangular half of the matrix between the two disease states for (max_distance, n_permutation)=(20,10).

A similar neighborship analysis was performed on the coarser cell-type level separately for the three malignant regions 6, 8, and 11, using TACCO's "co_occurrence_matrix" function with max_distance=20 and n_permutation=10.

Cell-type co-occurrence in Slide-seq data

Cell-type compositions relative to a spatial landmark, Region 2=muscularis, were evaluated using TACCO's "annotation_coordinate" function with max_distance=1000 and delta_distance=10. To reduce tissue structure bias from the muscularis, the distance dependency of cell-type frequency relations was evaluated only for beads deep in the "epithelial domain", defined as follows. The effective distance from stromal annotation was computed using TACCO's "annotation_coordinate" function (with max_distance=100, delta_distance=10, critical_neighbourhood_size=4.0) and only beads with a distance of at least 75µm were used. On these remaining beads, TACCO's "co_occurrence" function was used (with delta_distance=20, max_distance=1000) to compute cell types co-occurrence as a function of their distance.

Epithelial program neighborships in Slide-seq data

As for cell types above, neighborship relations were evaluated for epithelial programs in the AV Slide-seq samples using TACCO's "co_occurrence_matrix" function with max_distance=20 and

n_permutation=10, after selecting only the programs which make up at least 1% of the UMIs in the AV Slide-seq samples.

Ligand-Receptor analysis and spatially informed enrichment

For the ligand-receptor analysis on slide-seq data we used COMMOT v0.03[66]. COMMOT employs optimal transport to construct sender and receiver side receptor-ligand interactions for every bead in one run of COMMOT. After filtering to beads with at least 100 counts, we applied basic preprocessing (normalization, log1p-transformation) and load the CellChat database as done in the COMMOT tutorial (https://commot.readthedocs.io/en/latest/notebooks/Basic_usage.html). We then follow the Slide-Seq v2 analysis from the COMMOT paper ("slideseqv2-mouse-hippocampus/1-lr_signaling.ipynb" from https://doi.org/10.5281/zenodo.7272562) to filter the database and to reconstruct the spatial communication network on ligand-receptor pair and pathway level separately for every puck. In particular, we use the distance cutoff of 200µm for inference of ligand-receptor interactions. The resulting bead-wise sender- and receiver communication values were then used for enrichment analysis between disease states and between spatial regions.

Unlike the downstream Slide-Seq v2 analysis from the COMMOT paper ("slideseqv2-mouse-hippocampus/2-downstream_analysis.ipynb" from https://doi.org/10.5281/zenodo.7272562), we do not treat each bead on the puck as statistically independent observation in statistical tests, which leads to unreliably small p-values. Instead, we split all pucks along their axis of greatest extent (defined by the first principal component axis of the distribution of the spatial measurements) into

1010 spatial patches discarding 400µm of boundary layer (twice the COMMOT distance cutoff, also co-
1011 occurrences have decayed strongly at this distance) in between the patches to reduce the correlation
1012 between the patches. This is done iteratively with the patches to get a set of weakly correlated
1013 patches. On these patches we calculate the mean of the communication values and treat them as
1014 statistically independent observations for statistical tests. We argue that multiple sufficiently
1015 separated spatial patches of single spatial samples can be seen as multiple spatial samples using a
1016 spatial method with a smaller measurement area, and therefore can be treated as replicates. Unlike
1017 splitting into patches without removing a boundary layer, this procedure does not converge to the
1018 case of treating each bead as an independent observation as the number of iterations rises as it
1019 accounts for the spatial correlations between adjacent measurements. This gives a natural lower
1020 limit of p-values reachable with p-values rising again if too many splits are performed as too much
1021 data is lost to remove the correlations (**Supplementary item 4C,D**). We chose a number of
1022 iterations of 2 as a compromise between having more patches and not discarding much data. For
1023 the enrichment of the pathway communication values we used a two-sided Mann-Whitney U test
1024 across the patches and cite Benjamini-Hochberg FDR values. For the enrichment testing in a given
1025 group (e.g. an annotated spatial region) only those patches are used which have at least 100 beads
1026 on the patch.
1027
1028 Unless otherwise noted, for consistency, an analogous enrichment procedure was also used for
1029 COMMOT-unrelated quantities, like cell types, even though their spatial footprint is not as large
1030 as COMMOT's (distance cutoff of 200µm). For cluster-level cell type enrichment analyses on
1031 Slide-seq data, only the top 3 contributing subtypes were considered, to better reflect the expected
1032 properties of Slide-seq data and better compare with categorically annotated scRNA-seq data.

1033 **Mapping of mouse and human orthologs**

1034 We applied TACCO's functions "setup_orthology_converter" and "run_orthology_converter"
1035 with option "use_synonyms=True" to map human to mouse genes using the ortholog mapping
1036 from Mouse Genome Informatics (http://www.informatics.jax.org/homology.shtml). Specifically
1037 we used http://www.informatics.jax.org/downloads/reports/HOM_MouseHumanSequence.rpt
1038 (downloaded April 26th, 2021) for the analyses comparing T/NK compositions of human and
1039 mouse scRNAseq data, analyses comparing epithelial program similarity in human and mouse
1040 scRNAseq data, analyses involving the annotation of human scRNA-seq data with mouse regions,
1041 and analyses comparing the cell type and epithelial program correlations cross samples between
1042 human and mouse scRNA-seq data, and
1043 http://www.informatics.jax.org/downloads/reports/HOM_AllOrganism.rpt (downloaded on
1044 August 8th, 2022) for the analyses involving the scoring of mouse regions in mouse and human
1045 scRNA-seq data and leading to the CMS classification.

1046

1047 **GO term enrichment analysis**

1048 We used TACCO's functions "setup_goa_analysis" and "run_goa_analysis" to perform GO terms
1049 enrichment. As "gene_info_file" we used
1050 https://ftp.ncbi.nih.gov/gene/DATA/GENE_INFO/Mammalia/Mus_musculus.gene_info.gz, as
1051 "GO_obo_file" http://purl.obolibrary.org/obo/go/go-basic.obo, and as "gene2GO_file"
1052 https://ftp.ncbi.nih.gov/gene/DATA/gene2go.gz (all downloaded on August 10th, 2022).

1053

1054 **Comparison between mouse and human programs**

1055 To compare mouse and human epithelial expression programs[2], genes were mapped to mouse
1056 homologs using MGI homology information [subsection "Mapping of mouse and human
1057 orthologs"]. Mouse and human programs were then characterized by a single vector of mean
1058 expression per program in mouse gene space. Specifically, both mouse and human programs were
1059 defined such that their weighted sum approximates the expression profiles of the cells without any
1060 transformations. Programs and weights were normalized to sum to 1. To reduce batch-effects
1061 (including species-specific ones), a background expression profile was defined for each species
1062 dataset as the pseudo-bulk epithelial expression profile in the respective scRNAs-seq data.
1063 Program and background profiles were normalized to 10,000 counts and the log ratio of the
1064 normalized program and background expression vectors was used to define a vector for each
1065 species. Pearson correlation coefficients were calculated for each pair of program vectors (mouse
1066 vs. human).

1067 **Human expression program associations across mouse scRNA-seq**

1068 All sets of programs that were previously used to define human CRC tissue hubs[2] (epithelial, T/NK
1069 cells and myeloid cells) were mapped to mouse genes with MGI homology information [subsection
1070 "Mapping of mouse and human orthologs"] and then to mouse single cell data with TACCO, using
1071 TACCOs platform normalization to account for batch effects. The "annotate" function in TACCO
1072 was used with OT as the core annotation method, on the comparable subsets of cells from mouse
1073 and human single-cell datasets (*e.g.*, myeloid cells from mouse and human), with basic platform
1074 normalization, entropy regularization parameter epsilon 0.005, marginal relaxation parameter
1075 lambda of 0.1, and 4 iterations of bisectioning with a divisor of 3, and flat annotation prior
1076 distribution. The resulting probabilistic per-cell program annotations were aggregated to get

1077 probabilistic per-sample program annotations for all dysplastic mouse samples and CLR-
1078 transformed. For each pair of programs, the Pearson correlation coefficient was calculated on these
1079 transformed values.

1080

1081 **Annotating human scRNA-seq data with mouse-derived region information**

1082 Human scRNA-seq profiles[2] were mapped to mouse gene space using MGI homology information
1083 [subsection "Mapping of mouse and human orthologs"]. Working in the same expression space,
1084 the "annotate" function in TACCO with OT as core annotation method was used on the full human
1085 scRNA-seq and mouse Slide-seq dataset with basic platform normalization, entropy regularization
1086 parameter epsilon 0.005, marginal relaxation parameter lambda of 0.1, and 7 iterations of
1087 bisectioning with a divisor of 3, and 10-fold sub-clustering of the region annotations. The region
1088 transfer is done separately per compartment, with the Slide-seq compartment split as described
1089 above and the human scRNA-seq data split using the cell type annotation of the data. For
1090 validation, mapping was also performed with mouse scRNA-seq data, as well as mapping the
1091 region information from the mouse pucks back to themselves.

1092

1093 To test for enrichments of region annotations across disease state, region composition was
1094 aggregated to sample-level (for Slide-seq to 4-way split pucks), CLR-transformed, and enrichment
1095 was calculated using a two-sided Welch's t-test. This was done for region annotation on human
1096 and mouse scRNA-seq data, and on the original and mapped region annotation on the mouse Slide-
1097 seq data.

1098

1099 **Cell-type associations across samples**

1100 To compare associations of cell types across samples in human and mouse scRNA-seq the
1101 "clMidwayPr" cell type annotation in the human data[2] was aggregated to the same level as mouse
1102 cell type annotation, and then aggregated per sample and CLR-transformed. Pearson correlation
1103 coefficients were calculated for every cell type pair for different subsets of samples: all samples,
1104 normal samples, dysplastic, and for human MMRd/MMRp samples.

1105

1106 **Epithelial program associations across human samples**

1107 To determine epithelial program associations across human samples, TACCO's "annotate"
1108 function was used to annotate human epithelial scRNA-seq (after mapping to mouse orthologs
1109 using MGI homology information [subsection "Mapping of mouse and human orthologs"]) with
1110 mouse epithelial programs from mouse scRNA-seq data using OT as core method, basic platform
1111 normalization, entropy regularization parameter epsilon 0.005, marginal relaxation parameter
1112 lambda of 0.1, and 4 iterations of bisectioning with a divisor of 3. The remaining steps were
1113 performed as for cell-type association (subsection "Cell-type associations across samples").

1114

1115 **Scoring epithelial mouse regions in mouse and human epithelial pseudo-bulk data**

1116 The published processed and filtered count matrices were used (where available) or instead raw
1117 count matrices for single cell/nucleus RNA seq data from Pelka[2] (GEO accession number
1118 GSE178341; downloaded on August 8th, 2022), Chen[28] (Synapse IDs syn27056096,
1119 syn27056097, syn27056098, syn27056099; downloaded on August 5th, 2022), Khaliq[69] (GEO
1120 accession number GSE200997; downloaded on August 5th, 2022), Becker[29] (GEO accession
1121 number GSE201348; downloaded on August 5th, 2022), Zheng[68] (GEO accession number
1122 GSE161277; downloaded on August 5th, 2022; excluding 'blood' samples), Che[67] (GEO accession

1123 number GSE178318; downloaded on August 5th, 2022; only 'CRC' and 'LM' samples) and
1124 Joanito[70] (Synapse IDs syn26844072, syn26844073, syn26844078, syn26844087, syn26844111;
1125 downloaded on August 22nd, 2022; excluding the 'LymphNode' sample).

1126

1127 To subset the human single cell data to epithelial cells, the epithelial annotation was used where
1128 readily available[2,28]. For the remaining datasets[29,67–70], TACCOs tc.tl.annotate function was used
1129 with default parameters to transfer the 'cl295v11SubShort' annotation from Pelka[2], from which a
1130 compositional compartment annotation was constructed, and then a cell was assigned to the
1131 epithelial compartment if it had more than 95% epithelial fraction.

1132 To correct for batch effects between the different data sources, first batches were defined by
1133 species times protokoll: 'mouse-10x3p', 'mouse-SlideSeq', 'human-10x3p'[2,67,68,70], 'human-
1134 10x5p'[69,70], 'human-inDrop'[28], and 'human-snRNA'[29]. Then TACCO's "tc.pp.normalize_platform"
1135 function was used to determine per gene batch normalization factors using only the normal samples
1136 of one data source per batch (choosing the normal samples from Zheng for 'human-10x3p' and the
1137 normal 5' samples from Joanito for 'human-10x5p'). The resulting factors are then used to rescale
1138 the sample-by-gene count matrices for the full dataset per batch, i.e. including non-normal
1139 samples. The normalization factors are calculated with respect to an (arbitrarily chosen) normal
1140 reference dataset[68].

1141 The epithelial mouse region score was defined as the mean of the CLR-transformed expression
1142 values in the pseudo-bulk expression profile of the epithelial part of a dataset using the top 200
1143 differentially expressed genes between all regions by a one-sided Fisher's exact test.

1144 To account for species-specific biases (in-set *vs*. out-of-set prediction: the DEGs are calculated in
1145 mouse), the scores per region across samples were zero-centered and scaled to unit variance across

1146 all samples (including normal and non-normal samples and all batches) per species. A Principal
1147 Components Analysis (PCA) of the region scores across all species, batches and samples was
1148 conducted and the values for the first PC were compared between different conditions using a two-
1149 sided Mann-Whitney U test with Benjamini-Hochberg FDR.

1150

1151 **Assessing the relationship between mouse regions and CMS tumor classification**

1152 We used the package 'CMSclassifier' (https://github.com/Sage-Bionetworks/CMSclassifier)
1153 referred to in Ref.[71] to classify human pseudo-bulk CRC profiles from all samples which were
1154 not normal or unaffected from the human studies above into CMS classes. We determine the
1155 enrichment (Benjamini-Hochberg FDR, two-sided Welch's t test) of the same mouse region
1156 scores in the CMS classes.

1157

1158 **Assessing the relation between mouse regions and clinical endpoints in human bulk RNA-**
1159 **seq**

1160 Published RNA-seq data from the COAD and READ cohorts of TCGA PanCancerAtlas[95] were
1161 used. Mouse region scores were defined as the mean of the log1p-transformed, zero-centered and
1162 scaled expression values in the bulk expression profile using the top 200 differentially expressed
1163 genes between the malignant mouse regions (6, 8, 11) by a one-sided Fisher's Exact test
1164 (comparing each of the three regions to the other two). Scores were stratified into quartiles. PFI
1165 and OS was compared between patients with tumors whose scores were in the lowest and highest
1166 quartiles using the Logrank test as implemented in the lifelines package[96], followed by Benjamini-
1167 Hochberg FDR.

1168

1169 **Compositional enrichment analyses**

1170 Enrichments on compositional data (cell type compositions, etc.) were evaluated with a two-sided
1171 Welch's t test on sample level using CLR-transformed compositions followed by Benjamini-
1172 Hochberg FDR. For the enrichment of tdTomato, counts and ALR-transformation were used
1173 instead with all non-tdTomato counts used as reference compartment. Enrichment analyses were
1174 performed using TACCO's "enrichments" function.

1175

1176 **Code availability**

1177 The analysis code is available on GitHub (https://github.com/simonwm/mouseCRC).

1178 **Data availability**

1179 All data generated in this project is deposited on the Single Cell Portal and is available under the
1180 accession SCP1891 (https://singlecell.broadinstitute.org/single_cell/study/SCP1891). The raw
1181 scRNA-seq and Slide-seq data is also deposited on GEO under the GSE260801
1182 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE260801). The public human scRNA-
1183 seq datasets can be found on GEO under accession numbers GSE178341, GSE200997,
1184 GSE201348, GSE161277, GSE178318 and on Synapse under the IDs syn27056096,
1185 syn27056097, syn27056098, syn27056099, syn26844072, syn26844073, syn26844078,
1186 syn26844087, syn26844111.

1187 **ACKNOWLEDGEMENTS**

1188 We thank N. Friedman, I. Benhar, N. Habib, M. Biton, K. Geiger-Schuller, J.C. Hütter, B.
1189 Dumitrascu, E. Baker and A. Greenwald for helpful discussions. We thank C. McCabe, O.
1190 Kuksenko and I. Barrera for technical assistance. We thank P. Yadollahpour, E. Dhaval, G. Smith-
1191 Rosario, S. Vickovic, D. Schapiro, S. Farhi, D. Abbondanza, A. Segerstolpe and T. Biancalani for
1192 their important contribution to this project. We thank L. Gaffney and A. Hupalowska for help with
1193 figure preparation. S.M. was supported by a DFG research fellowship (MA 9108/1-1), J.K. was
1194 supported by a HFSP long term fellowship (LT000452/2019-L), A.R. was a Howard Hughes
1195 Medical Institute (HHMI) Investigator when conducting this work. Work was supported by the
1196 Klarman Cell Observatory, a CEGS grant (5RM1HG006193-09) from the NHGRI, the
1197 NIH/NIAID (grants 1U24 CA180922, 1U19 MH114821, 1RC2 DK114784), the MIT Ludwig
1198 Center, the Manton Family Foundation, and HHMI (A.R.); Azrieli Foundation Early Career
1199 Faculty Fellowship, and an ISF Research Grant (1079/21) (M.N.), the Center for Interdisciplinary
1200 Data Science Research at the Hebrew University of Jerusalem (N.M. and M.N.), the Israeli Council
1201 for Higher Education Ph.D. fellowship (N.M.), SU2C Peggy Prescott Early Career Scientist Award
1202 PA-6146, SU2C Phillip A. Sharp Award SU2C-AACR-PS-32 and NIH/NCI R00CA259511
1203 (K.P.), NIH/NCI R01 CA208756; Arthur, Sandra, and Sarah Irving Fund for Gastrointestinal
1204 Immuno-Oncology (N.H.), NIH/NCI R01CA257523, MIT Stem Cell Initiative (Foundation MIT)
1205 (O.Y.), and NIH R37CA259363, R21CA256414, R21DK125911, R41EB032693, R01CA254108,
1206 R01CA256530, and R01CA244359; DOD W81XWH-20-1-0203; and a Duke-NC State
1207 Translational Research Grant (J.R.). The authors gratefully acknowledge LMU Klinikum for
1208 providing computing resources on their Clinical Open Research Engine (CORE) and the

Bioinformatic Core Facility of the Biomedical Center Munich for providing computing resources on their HPC system.

AUTHOR CONTRIBUTIONS

I.A-D. and A.R. conceived the study and designed experiments. I.A-D. J.R., S.Y., E.M., T.D., L.C. and D.D. performed experiments, with guidance from O.Y. and A.R.. S.M., J.K., and N.M. developed computational approaches and analyzed data with I.A-D., M.H., E.M. and R.S, and with guidance from M.N. and A.R.. M.H., E.M., J.C., K.P., A.M. and G.M.B. generated and analyzed human data, with guidance from N.H., I.T., N.H., F.C., O.Y., O.R-R, and A.R.. A.N. and H.X. generated and processed the multiplex *in-situ* RNA data, with guidance from J.L. and M.R.. M.N and A.R provided supervision and acquired funding. I.A-D., S.M., M.N. and A.R. wrote the manuscript with input from all authors.

DECLARATION OF INTERESTS

A.M. has served a consultant/advisory role for Third Rock Ventures, Asher Biotherapeutics, Abata Therapeutics, ManaT Bio, Flare Therapeutics, venBio Partners, BioNTech, Rheos Medicines and Checkmate Pharmaceuticals, is currently a part-time Entrepreneur in Residence at Third Rock Ventures, is an equity holder in ManaT Bio, Asher Biotherapeutics and Abata Therapeutics, and has received research funding support from Bristol-Myers Squibb. G.M.B. has sponsored research agreements with InterVenn Biosciences, Palleon Pharmaceuticals, Olink Proteomics, and Teiko Bio. G.M.B. is a consultant for Ankyra Therapeutics and InterVenn Bio. G.M.B. has been on scientific advisory boards for Merck, Iovance, Nektar Therapeutics, Instil Bio, and Ankyra Therapeutics. G.M.B. holds equity in Ankyra Therapeutics. J.L. and M.R. hold equity in 10xGenomics. N.H. holds equity in BioNTech and is a founder of Related Sciences/DangerBio.

F.C. is a founder and holds equity in Curio Biosciences. O.Y. holds equity and is a SAB member of AVA Lifesciences. A.R. and O.R.-R. are co-inventors on patent applications filed by the Broad Institute for inventions related to single cell genomics. O.R.-R. has given numerous lectures on the subject of single cell genomics to a wide variety of audiences and in some cases, has received remuneration to cover time and costs. O.R.-R. is an employee of Genentech since October 19, 2020 and has equity in Roche. A.R. is a co-founder and equity holder of Celsius Therapeutics, an equity holder in Immunitas, and was an SAB member of ThermoFisher Scientific, Syros Pharmaceuticals, Neogene Therapeutics and Asimov until July 31, 2020. From August 1, 2020, A.R. is an employee of Genentech and has equity in Roche.

1241 **References**

1242 1. McAllister, S.S., and Weinberg, R.A. (2014). The tumour-induced systemic environment as
1243 a critical regulator of cancer progression and metastasis. Nat. Cell Biol. *16*, 717–727.

1244 2. Pelka, K., Hofree, M., Chen, J.H., Sarkizova, S., Pirl, J.D., Jorgji, V., Bejnood, A., Dionne,
1245 D., Ge, W.H., Xu, K.H., et al. (2021). Spatially organized multicellular immune hubs in
1246 human colorectal cancer. Cell *184*, 4734-4752.e20.

1247 3. Wagner, J., Rapsomaniki, M.A., Chevrier, S., Anzeneder, T., Langwieder, C., Dykgers, A.,
1248 Rees, M., Ramaswamy, A., Muenst, S., Soysal, S.D., et al. (2019). A Single-Cell Atlas of the
1249 Tumor and Immune Ecosystem of Human Breast Cancer. Cell *177*, 1330-1345.e18.

1250 4. Hunter, M.V., Moncada, R., Weiss, J.M., Yanai, I., and White, R.M. (2021). Spatially
1251 resolved transcriptomics reveals the architecture of the tumor-microenvironment interface.
1252 Nat. Commun. *12*, 1–16.

1253 5. Keren, L., Bosse, M., Marquez, D., Angoshtari, R., Jain, S., Varma, S., Yang, S.-R., Kurian,
1254 A., Van Valen, D., West, R., et al. (2018). A Structured Tumor-Immune Microenvironment
1255 in Triple Negative Breast Cancer Revealed by Multiplexed Ion Beam Imaging. Cell *174*,
1256 1373-1387.e19.

1257 6. Schürch, C.M., Bhate, S.S., Barlow, G.L., Phillips, D.J., Noti, L., Zlobec, I., Chu, P., Black,
1258 S., Demeter, J., McIlwain, D.R., et al. (2020). Coordinated cellular neighborhoods orchestrate
1259 antitumoral immunity at the colorectal cancer invasive front. Cell *182*, 1341-1359.e19.

1260 7. Jackson, H.W., Fischer, J.R., Zanotelli, V.R.T., Ali, H.R., Mechera, R., Soysal, S.D., Moch,
1261 H., Muenst, S., Varga, Z., Weber, W.P., et al. (2020). The single-cell pathology landscape of
1262 breast cancer. Nature *578*, 615–620.

1263 8. Jerby-Arnon, L., Shah, P., Cuoco, M.S., Rodman, C., Su, M.-J., Melms, J.C., Leeson, R.,
1264 Kanodia, A., Mei, S., Lin, J.-R., et al. (2018). A Cancer Cell Program Promotes T Cell
1265 Exclusion and Resistance to Checkpoint Blockade. Cell *175*, 984-997.e24.

1266 9. Grünwald, B.T., Devisme, A., Andrieux, G., Vyas, F., Aliar, K., McCloskey, C.W., Macklin,
1267 A., Jang, G.H., Denroche, R., Romero, J.M., et al. (2021). Spatially confined sub-tumor
1268 microenvironments in pancreatic cancer. Cell *184*, 5577-5592.e18.

1269 10. Goltsev, Y., Samusik, N., Kennedy-Darling, J., Bhate, S., Hale, M., Vazquez, G., Black, S.,
1270 and Nolan, G.P. (2018). Deep Profiling of Mouse Splenic Architecture with CODEX
1271 Multiplexed Imaging. Cell *174*, 968-981.e15.

1272 11. Angelo, M., Bendall, S.C., Finck, R., Hale, M.B., Hitzman, C., Borowsky, A.D., Levenson,
1273 R.M., Lowe, J.B., Liu, S.D., Zhao, S., et al. (2014). Multiplexed ion beam imaging of human
1274 breast tumors. Nat. Med. *20*, 436–442.

1275 12. Giesen, C., Wang, H.A.O., Schapiro, D., Zivanovic, N., Jacobs, A., Hattendorf, B., Schüffler,
1276 P.J., Grolimund, D., Buhmann, J.M., Brandt, S., et al. (2014). Highly multiplexed imaging of
1277 tumor tissues with subcellular resolution by mass cytometry. Nat. Methods *11*, 417–422.

1278 13. Ståhl, P.L., Salmén, F., Vickovic, S., Lundmark, A., Navarro, J.F., Magnusson, J.,
1279 Giacomello, S., Asp, M., Westholm, J.O., Huss, M., et al. (2016). Visualization and analysis
1280 of gene expression in tissue sections by spatial transcriptomics. Science *353*, 78–82.

1281 14. Rodriques, S.G., Stickels, R.R., Goeva, A., Martin, C.A., Murray, E., Vanderburg, C.R.,
1282 Welch, J., Chen, L.M., Chen, F., and Macosko, E.Z. (2019). Slide-seq: A scalable technology
1283 for measuring genome-wide expression at high spatial resolution. Science *363*, 1463–1467.

1284 15. Stickels, R.R., Murray, E., Kumar, P., Li, J., Marshall, J.L., Di Bella, D.J., Arlotta, P.,
1285 Macosko, E.Z., and Chen, F. (2021). Highly sensitive spatial transcriptomics at near-cellular
1286 resolution with Slide-seqV2. Nat. Biotechnol. *39*, 313–319.

1287 16. Marx, V. (2021). Publisher Correction: Method of the Year: spatially resolved
1288 transcriptomics. Nat. Methods *18*, 219.

1289 17. Waylen, L.N., Nim, H.T., Martelotto, L.G., and Ramialison, M. (2020). From whole-mount
1290 to single-cell spatial assessment of gene expression in 3D. Commun Biol *3*, 602.

1291 18. Palla, G., Fischer, D.S., Regev, A., and Theis, F.J. (2022). Spatial components of molecular
1292 tissue biology. Nat. Biotechnol. https://doi.org/10.1038/s41587-021-01182-1.

1293 19. Cable, D.M., Murray, E., Zou, L.S., Goeva, A., Macosko, E.Z., Chen, F., and Irizarry, R.A.
1294 (2021). Robust decomposition of cell type mixtures in spatial transcriptomics. Nat.
1295 Biotechnol., 1–10.

1296 20. Kleshchevnikov, V., Shmatko, A., Dann, E., Aivazidis, A., King, H.W., Li, T., Elmentaite,
1297 R., Lomakin, A., Kedlian, V., Gayoso, A., et al. (2022). Cell2location maps fine-grained cell
1298 types in spatial transcriptomics. Nat. Biotechnol. *40*, 661–671.

1299 21. Mages, S., Moriel, N., Avraham-Davidi, I., Murray, E., Watter, J., Chen, F., Rozenblatt-
1300 Rosen, O., Klughammer, J., Regev, A., and Nitzan, M. (2023). TACCO unifies annotation
1301 transfer and decomposition of cell identities for single-cell and spatial omics. Nat.
1302 Biotechnol., 1–9.

1303 22. Lopez, R., Li, B., Keren-Shaul, H., Boyeau, P., Kedmi, M., Pilzer, D., Jelinski, A., Yofe, I.,
1304 David, E., Wagner, A., et al. (2022). DestVI identifies continuums of cell types in spatial
1305 transcriptomics data. Nat. Biotechnol. *40*, 1360–1369.

1306 23. Biancalani, T., Scalia, G., Buffoni, L., Avasthi, R., Lu, Z., Sanger, A., Tokcan, N.,
1307 Vanderburg, C.R., Segerstolpe, Å., Zhang, M., et al. (2021). Deep learning and alignment of
1308 spatially resolved single-cell transcriptomes with Tangram. Nat. Methods *18*, 1352–1362.

1309 24. Cancer Genome Atlas Network (2012). Comprehensive molecular characterization of human
1310 colon and rectal cancer. Nature *487*, 330–337.

1311 25. Kwong, L.N., and Dove, W.F. (2009). APC and its modifiers in colon cancer. Adv. Exp. Med.
1312 Biol. *656*, 85–106.

1313 26. Fearon, E.R. (2011). Molecular genetics of colorectal cancer. Annu. Rev. Pathol. *6*, 479–507.

1314 27. Fearon, E.R., and Vogelstein, B. (1990). A genetic model for colorectal tumorigenesis. Cell
1315 *61*, 759–767.

1316 28. Chen, B., Scurrah, C.R., McKinley, E.T., Simmons, A.J., Ramirez-Solano, M.A., Zhu, X.,
1317 Markham, N.O., Heiser, C.N., Vega, P.N., Rolong, A., et al. (2021). Differential pre-
1318 malignant programs and microenvironment chart distinct paths to malignancy in human
1319 colorectal polyps. Cell *184*, 6262-6280.e26.

1320 29. Becker, W.R., Nevins, S.A., Chen, D.C., Chiu, R., Horning, A.M., Guha, T.K., Laquindanum,
1321 R., Mills, M., Chaib, H., Ladabaum, U., et al. (2022). Single-cell analyses define a continuum
1322 of cell state and composition changes in the malignant transformation of polyps to colorectal
1323 cancer. Nat. Genet., 1–11.

1324 30. Roper, J., Tammela, T., Cetinbas, N.M., Akkad, A., Roghanian, A., Rickelt, S., Almeqdadi,
1325 M., Wu, K., Oberli, M.A., Sánchez-Rivera, F.J., et al. (2017). In vivo genome editing and
1326 organoid transplantation models of colorectal cancer and metastasis. Nat. Biotechnol. *35*,
1327 569–576.

1328 31. Roper, J., Tammela, T., Akkad, A., Almeqdadi, M., Santos, S.B., Jacks, T., and Yilmaz, Ö.H.
1329 (2018). Colonoscopy-based colorectal cancer modeling in mice with CRISPR–Cas9 genome
1330 editing and organoid transplantation. Nat. Protoc. *13*, 217–234.

1331 32. Golovko, D., Kedrin, D., Yilmaz, Ö.H., and Roper, J. (2015). Colorectal cancer models for
1332 novel drug discovery. Expert Opin. Drug Discov. *10*, 1217–1229.

1333 33. Gyllborg, D., Langseth, C.M., Qian, X., Choi, E., Salas, S.M., Hilscher, M.M., Lein, E.S.,
1334 and Nilsson, M. (2020). Hybridization-based in situ sequencing (HybISS) for spatially
1335 resolved transcriptomics in human and mouse brain tissue. Nucleic Acids Res. *48*, e112.

1336 34. Hagerling, C., Gonzalez, H., Salari, K., Wang, C.-Y., Lin, C., Robles, I., van Gogh, M.,
1337 Dejmek, A., Jirström, K., and Werb, Z. (2019). Immune effector monocyte–neutrophil
1338 cooperation induced by the primary tumor prevents metastatic progression of breast cancer.
1339 Proc. Natl. Acad. Sci. U. S. A. *116*, 21704–21714.

1340 35. Arenberg, D.A., Keane, M.P., DiGiovine, B., Kunkel, S.L., Strom, S.R., Burdick, M.D.,
1341 Iannettoni, M.D., and Strieter, R.M. (2000). Macrophage infiltration in human non-small-cell
1342 lung cancer: the role of CC chemokines. Cancer Immunol. Immunother. *49*, 63–70.

1343 36. Smith, P.M., and Garrett, W.S. (2011). The gut microbiota and mucosal T cells. Front.
1344 Microbiol. *2*, 111.

1345 37. La Manno, G., Soldatov, R., Zeisel, A., Braun, E., Hochgerner, H., Petukhov, V.,
1346 Lidschreiber, K., Kastriti, M.E., Lönnerberg, P., Furlan, A., et al. (2018). RNA velocity of
1347 single cells. Nature *560*, 494–498.

1348 38. Ziyad, S., and Iruela-Arispe, M.L. (2011). Molecular mechanisms of tumor angiogenesis.
1349 Genes Cancer *2*, 1085–1096.

1350 39. Zeng, Z.-S., Shu, W.-P., Cohen, A.M., and Guillem, J.G. (2002). Matrix metalloproteinase-7
1351 expression in colorectal cancer liver metastases: evidence for involvement of MMP-7
1352 activation in human cancer metastases. Clin. Cancer Res. *8*, 144–148.

1353 40. Li, D., Peng, Z., Tang, H., Wei, P., Kong, X., Yan, D., Huang, F., Li, Q., Le, X., Li, Q., et al.
1354 (2011). KLF4-mediated negative regulation of IFITM3 expression plays a critical role in
1355 colon cancer pathogenesis. Clin. Cancer Res. *17*, 3558–3568.

1356 41. Kitamura, T., Kometani, K., Hashida, H., Matsunaga, A., Miyoshi, H., Hosogi, H., Aoki, M.,
1357 Oshima, M., Hattori, M., Takabayashi, A., et al. (2007). SMAD4-deficient intestinal tumors
1358 recruit CCR1+ myeloid cells that promote invasion. Nat. Genet. *39*, 467–475.

1359 42. Haber, A.L., Biton, M., Rogel, N., Herbst, R.H., Shekhar, K., Smillie, C., Burgin, G., Delorey,
1360 T.M., Howitt, M.R., Katz, Y., et al. (2017). A single-cell survey of the small intestinal
1361 epithelium. Nature *551*, 333–339.

1362 43. Biton, M., Haber, A.L., Rogel, N., Burgin, G., Beyaz, S., Schnell, A., Ashenberg, O., Su, C.-
1363 W., Smillie, C., Shekhar, K., et al. (2018). T Helper Cell Cytokines Modulate Intestinal Stem
1364 Cell Renewal and Differentiation. Cell *175*, 1307-1320.e22.

1365 44. Smillie, C.S., Biton, M., Ordovas-Montanes, J., Sullivan, K.M., Burgin, G., Graham, D.B.,
1366 Herbst, R.H., Rogel, N., Slyper, M., Waldman, J., et al. (2019). Intra- and Inter-cellular
1367 Rewiring of the Human Colon during Ulcerative Colitis. Cell *178*, 714-730.e22.

1368 45. Welch, J.D., Kozareva, V., Ferreira, A., Vanderburg, C., Martin, C., and Macosko, E.Z.
1369 (2019). Single-Cell Multi-omic Integration Compares and Contrasts Features of Brain Cell
1370 Identity. Cell *177*, 1873-1887.e17.

1371 46. van Neerven, S.M., de Groot, N.E., Nijman, L.E., Scicluna, B.P., van Driel, M.S., Lecca,
1372 M.C., Warmerdam, D.O., Kakkar, V., Moreno, L.F., Vieira Braga, F.A., et al. (2021). Apc-
1373 mutant cells act as supercompetitors in intestinal tumour initiation. Nature *594*, 436–441.

1374 47. Flanagan, D.J., Pentinmikko, N., Luopajärvi, K., Willis, N.J., Gilroy, K., Raven, A.P.,
1375 Mcgarry, L., Englund, J.I., Webb, A.T., Scharaw, S., et al. (2021). NOTUM from Apc-mutant
1376 cells biases clonal competition to initiate cancer. Nature *594*, 430–435.

1377 48. Clevers, H. (2006). Wnt/beta-catenin signaling in development and disease. Cell *127*, 469–
1378 480.

1379 49. Folkman, J. (2002). Role of angiogenesis in tumor growth and metastasis. Semin. Oncol. *29*,
1380 15–18.

1381 50. Lasry, A., Zinger, A., and Ben-Neriah, Y. (2016). Inflammatory networks underlying
1382 colorectal cancer. Nat. Immunol. *17*, 230–240.

1383 51. Fleming, M., Ravula, S., Tatishchev, S.F., and Wang, H.L. (2012). Colorectal carcinoma:
1384 Pathologic aspects. J. Gastrointest. Oncol. *3*, 153–173.

1385 52. Winkler, J., Abisoye-Ogunniyan, A., Metcalf, K.J., and Werb, Z. (2020). Concepts of
1386 extracellular matrix remodelling in tumour progression and metastasis. Nat. Commun. *11*, 1–
1387 19.

1388 53. Drev, D., Harpain, F., Beer, A., Stift, A., Gruber, E.S., Klimpfinger, M., Thalhammer, S.,
1389 Reti, A., Kenner, L., Bergmann, M., et al. (2019). Impact of Fibroblast-Derived SPARC on
1390 Invasiveness of Colorectal Cancer Cells. Cancers *11*.
1391 https://doi.org/10.3390/cancers11101421.

1392 54. Burden, R.E., Gormley, J.A., Jaquin, T.J., Small, D.M., Quinn, D.J., Hegarty, S.M., Ward,
1393 C., Walker, B., Johnston, J.A., Olwill, S.A., et al. (2009). Antibody-mediated inhibition of
1394 cathepsin S blocks colorectal tumor invasion and angiogenesis. Clin. Cancer Res. *15*, 6042–
1395 6051.

1396 55. Basu, S., Cheriyamundath, S., Gavert, N., Brabletz, T., Haase, G., and Ben-Ze'ev, A. (2019).
1397 Increased expression of cathepsin D is required for L1-mediated colon cancer progression.
1398 Oncotarget *10*, 5217–5228.

1399 56. Spencer, J., and Sollid, L.M. (2016). The human intestinal B-cell response. Mucosal
1400 Immunol. *9*, 1113–1124.

1401 57. Sautès-Fridman, C., Petitprez, F., Calderaro, J., and Fridman, W.H. (2019). Tertiary lymphoid
1402 structures in the era of cancer immunotherapy. Nat. Rev. Cancer *19*, 307–325.

1403 58. Mendez, M.G., Kojima, S.-I., and Goldman, R.D. (2010). Vimentin induces changes in cell
1404 shape, motility, and adhesion during the epithelial to mesenchymal transition. FASEB J. *24*,
1405 1838–1851.

1406 59. Lu, M.-H., Huang, C.-C., Pan, M.-R., Chen, H.-H., and Hung, W.-C. (2012). Prospero
1407 homeobox 1 promotes epithelial-mesenchymal transition in colon cancer cells by inhibiting
1408 E-cadherin via miR-9. Clin. Cancer Res. *18*, 6416–6425.

1409 60. Oliemuller, E., Newman, R., Tsang, S.M., Foo, S., Muirhead, G., Noor, F., Haider, S.,
1410 Aurrekoetxea-Rodríguez, I., Vivanco, M. dM, and Howard, B.A. (2020). SOX11 promotes
1411 epithelial/mesenchymal hybrid state and alters tropism of invasive breast cancer cells. Elife
1412 *9*, e58374.

1413 61. Marjanovic, N.D., Hofree, M., Chan, J.E., Canner, D., Wu, K., Trakala, M., Hartmann, G.G.,
1414 Smith, O.C., Kim, J.Y., Evans, K.V., et al. (2020). Emergence of a High-Plasticity Cell State
1415 during Lung Cancer Evolution. Cancer Cell *38*, 229-246.e13.

1416 62. Puram, S.V., Tirosh, I., Parikh, A.S., Patel, A.P., Yizhak, K., Gillespie, S., Rodman, C., Luo,
1417 C.L., Mroz, E.A., Emerick, K.S., et al. (2017). Single-Cell Transcriptomic Analysis of
1418 Primary and Metastatic Tumor Ecosystems in Head and Neck Cancer. Cell *171*, 1611-
1419 1624.e24.

1420 63. Schwab, R.H.M., Amin, N., Flanagan, D.J., Johanson, T.M., Phesse, T.J., and Vincan, E.
1421 (2018). Wnt is necessary for mesenchymal to epithelial transition in colorectal cancer cells.
1422 Dev. Dyn. *247*, 521–530.

1423 64. DiMeo, T.A., Anderson, K., Phadke, P., Fan, C., Perou, C.M., Naber, S., and Kuperwasser,
1424 C. (2009). A novel lung metastasis signature links Wnt signaling with cancer cell self-renewal
1425 and epithelial-mesenchymal transition in basal-like breast cancer. Cancer Res. *69*, 5364–
1426 5373.

1427 65. Penninger, J.M., and Crabtree, G.R. (1999). The Actin Cytoskeleton and Lymphocyte
1428 Activation. Cell *96*, 9–12.

1429 66. Cang, Z., Zhao, Y., Almet, A.A., Stabell, A., Ramos, R., Plikus, M.V., Atwood, S.X., and
1430 Nie, Q. (2023). Screening cell–cell communication in spatial transcriptomics via collective
1431 optimal transport. Nat. Methods *20*, 218–228.

1432 67. Che, L.-H., Liu, J.-W., Huo, J.-P., Luo, R., Xu, R.-M., He, C., Li, Y.-Q., Zhou, A.-J., Huang,
1433 P., Chen, Y.-Y., et al. (2021). A single-cell atlas of liver metastases of colorectal cancer
1434 reveals reprogramming of the tumor microenvironment in response to preoperative
1435 chemotherapy. Cell Discovery *7*, 1–21.

1436 68. Zheng, X., Song, J., Yu, C., Zhou, Z., Liu, X., Yu, J., Xu, G., Yang, J., He, X., Bai, X., et al.
1437 (2022). Single-cell transcriptomic profiling unravels the adenoma-initiation role of protein
1438 tyrosine kinases during colorectal tumorigenesis. Signal Transduction and Targeted Therapy
1439 *7*, 1–14.

1440 69. Khaliq, A.M., Erdogan, C., Kurt, Z., Turgut, S.S., Grunvald, M.W., Rand, T., Khare, S.,
1441 Borgia, J.A., Hayden, D.M., Pappas, S.G., et al. (2022). Correction: Refining colorectal
1442 cancer classification and clinical stratification through a single-cell atlas. Genome Biol. *23*,
1443 156.

1444 70. Joanito, I., Wirapati, P., Zhao, N., Nawaz, Z., Yeo, G., Lee, F., Eng, C.L.P., Macalinao, D.C.,
1445 Kahraman, M., Srinivasan, H., et al. (2022). Single-cell and bulk transcriptome sequencing
1446 identifies two epithelial tumor cell states and refines the consensus molecular classification
1447 of colorectal cancer. Nat. Genet. *54*, 963–975.

1448 71. Guinney, J., Dienstmann, R., Wang, X., de Reyniès, A., Schlicker, A., Soneson, C., Marisa,
1449 L., Roepman, P., Nyamundanda, G., Angelino, P., et al. (2015). The consensus molecular
1450 subtypes of colorectal cancer. Nat. Med. *21*, 1350–1356.

1451 72. West, N.R., Hegazy, A.N., Owens, B.M.J., Bullers, S.J., Linggi, B., Buonocore, S., Coccia,
1452 M., Görtz, D., This, S., Stockenhuber, K., et al. (2017). Oncostatin M drives intestinal

1453 inflammation and predicts response to tumor necrosis factor-neutralizing therapy in patients
1454 with inflammatory bowel disease. Nat. Med. *23*, 579–589.

1455 73. West, N.R., Murray, J.I., and Watson, P.H. (2013). Oncostatin-M promotes phenotypic
1456 changes associated with mesenchymal and stem cell-like differentiation in breast cancer.
1457 Oncogene *33*, 1485–1494.

1458 74. Smigiel, J.M., Parameswaran, N., and Jackson, M.W. (2017). Potent EMT and CSC
1459 phenotypes are induced by oncostatin-M in pancreatic cancer. Mol. Cancer Res. *15*, 478–488.

1460 75. Hara, T., Chanoch-Myers, R., Mathewson, N.D., Myskiw, C., Atta, L., Bussema, L.,
1461 Eichhorn, S.W., Greenwald, A.C., Kinker, G.S., Rodman, C., et al. (2021). Interactions
1462 between cancer cells and immune cells drive transitions to mesenchymal-like states in
1463 glioblastoma. Cancer Cell *39*, 779-792.e11.

1464 76. de Sousa e Melo, F., Kurtova, A.V., Harnoss, J.M., Kljavin, N., Hoeck, J.D., Hung, J.,
1465 Anderson, J.E., Storm, E.E., Modrusan, Z., Koeppen, H., et al. (2017). A distinct role for
1466 Lgr5+ stem cells in primary and metastatic colon cancer. Nature *543*, 676–680.

1467 77. Schepers, A.G., Snippert, H.J., Stange, D.E., van den Born, M., van Es, J.H., van de Wetering,
1468 M., and Clevers, H. (2012). Lineage tracing reveals Lgr5+ stem cell activity in mouse
1469 intestinal adenomas. Science *337*, 730–735.

1470 78. André, T., Shiu, K.-K., Kim, T.W., Jensen, B.V., Jensen, L.H., Punt, C., Smith, D., Garcia-
1471 Carbonero, R., Benavides, M., Gibbs, P., et al. (2020). Pembrolizumab in Microsatellite-
1472 Instability–High Advanced Colorectal Cancer. N. Engl. J. Med. *383*, 2207–2218.

1473 79. Kuraguchi, M., Wang, X.-P., Bronson, R.T., Rothenberg, R., Ohene-Baah, N.Y., Lund, J.J.,
1474 Kucherlapati, M., Maas, R.L., and Kucherlapati, R. (2006). Adenomatous polyposis coli
1475 (APC) is required for normal development of skin and thymus. PLoS Genet. *2*, e146.

1476 80. Johnson, L., Mercer, K., Greenbaum, D., Bronson, R.T., Crowley, D., Tuveson, D.A., and
1477 Jacks, T. (2001). Somatic activation of the K-ras oncogene causes early onset lung cancer in
1478 mice. Nature *410*, 1111–1116.

1479 81. Madisen, L., Zwingman, T.A., Sunkin, S.M., Oh, S.W., Zariwala, H.A., Gu, H., Ng, L.L.,
1480 Palmiter, R.D., Hawrylycz, M.J., Jones, A.R., et al. (2009). A robust and high-throughput Cre
1481 reporting and characterization system for the whole mouse brain. Nat. Neurosci. *13*, 133–140.

1482 82. Marino, S., Vooijs, M., van Der Gulden, H., Jonkers, J., and Berns, A. (2000). Induction of
1483 medulloblastomas in p53-null mutant mice by somatic inactivation of Rb in the external
1484 granular layer cells of the cerebellum. Genes Dev. *14*, 994–1004.

1485 83. el Marjou, F., Janssen, K.-P., Chang, B.H.-J., Li, M., Hindie, V., Chan, L., Louvard, D.,
1486 Chambon, P., Metzger, D., and Robine, S. (2004). Tissue-specific and inducible Cre-
1487 mediated recombination in the gut epithelium. Genesis *39*, 186–193.

1488 84. Stoeckius, M., Zheng, S., Houck-Loomis, B., Hao, S., Yeung, B.Z., Mauck, W.M., 3rd,
1489 Smibert, P., and Satija, R. (2018). Cell Hashing with barcoded antibodies enables
1490 multiplexing and doublet detection for single cell genomics. Genome Biol. *19*, 224.

1491 85. Qian, X., Harris, K.D., Hauling, T., Nicoloutsopoulos, D., Muñoz-Manchado, A.B., Skene,
1492 N., Hjerling-Leffler, J., and Nilsson, M. (2019). Probabilistic cell typing enables fine mapping
1493 of closely related cell types in situ. Nat. Methods *17*, 101–106.

1494 86. Li, B., Gould, J., Yang, Y., Sarkizova, S., Tabaka, M., Ashenberg, O., Rosen, Y., Slyper, M.,
1495 Kowalczyk, M.S., Villani, A.-C., et al. (2020). Cumulus provides cloud-based data analysis
1496 for large-scale single-cell and single-nucleus RNA-seq. Nat. Methods *17*, 793–798.

1497 87. Wolf, F.A., Angerer, P., and Theis, F.J. (2018). SCANPY: large-scale single-cell gene
1498 expression data analysis. Genome Biol. *19*, 15.

1499 88. Wolock, S.L., Lopez, R., and Klein, A.M. (2019). Scrublet: Computational Identification of
1500 Cell Doublets in Single-Cell Transcriptomic Data. Cell Syst *8*, 281-291.e9.

1501 89. Aran, D., Looney, A.P., Liu, L., Wu, E., Fong, V., Hsu, A., Chak, S., Naikawadi, R.P.,
1502 Wolters, P.J., Abate, A.R., et al. (2019). Reference-based analysis of lung single-cell
1503 sequencing reveals a transitional profibrotic macrophage. Nat. Immunol. *20*, 163–172.

1504 90. Tabula Muris Consortium, Overall coordination, Logistical coordination, Organ collection
1505 and processing, Library preparation and sequencing, Computational data analysis, Cell type
1506 annotation, Writing group, Supplemental text writing group, and Principal investigators
1507 (2018). Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. Nature *562*,
1508 367–372.

1509 91. Pedregosa, F., Varoquaux, G., Gramfort, A., and Michel, V. Scikit-learn: Machine Learning
1510 in Python. 12 (85): 2825–2830. : http://jmlr. org/papers/v12 ….

1511 92. Polański, K., Young, M.D., Miao, Z., Meyer, K.B., Teichmann, S.A., and Park, J.-E. (2020).
1512 BBKNN: fast batch alignment of single cell transcriptomes. Bioinformatics *36*, 964–965.

1513 93. Rao, J.N., and Wang, J.-Y. (2010). Intestinal Architecture and Development (Morgan &
1514 Claypool Life Sciences).

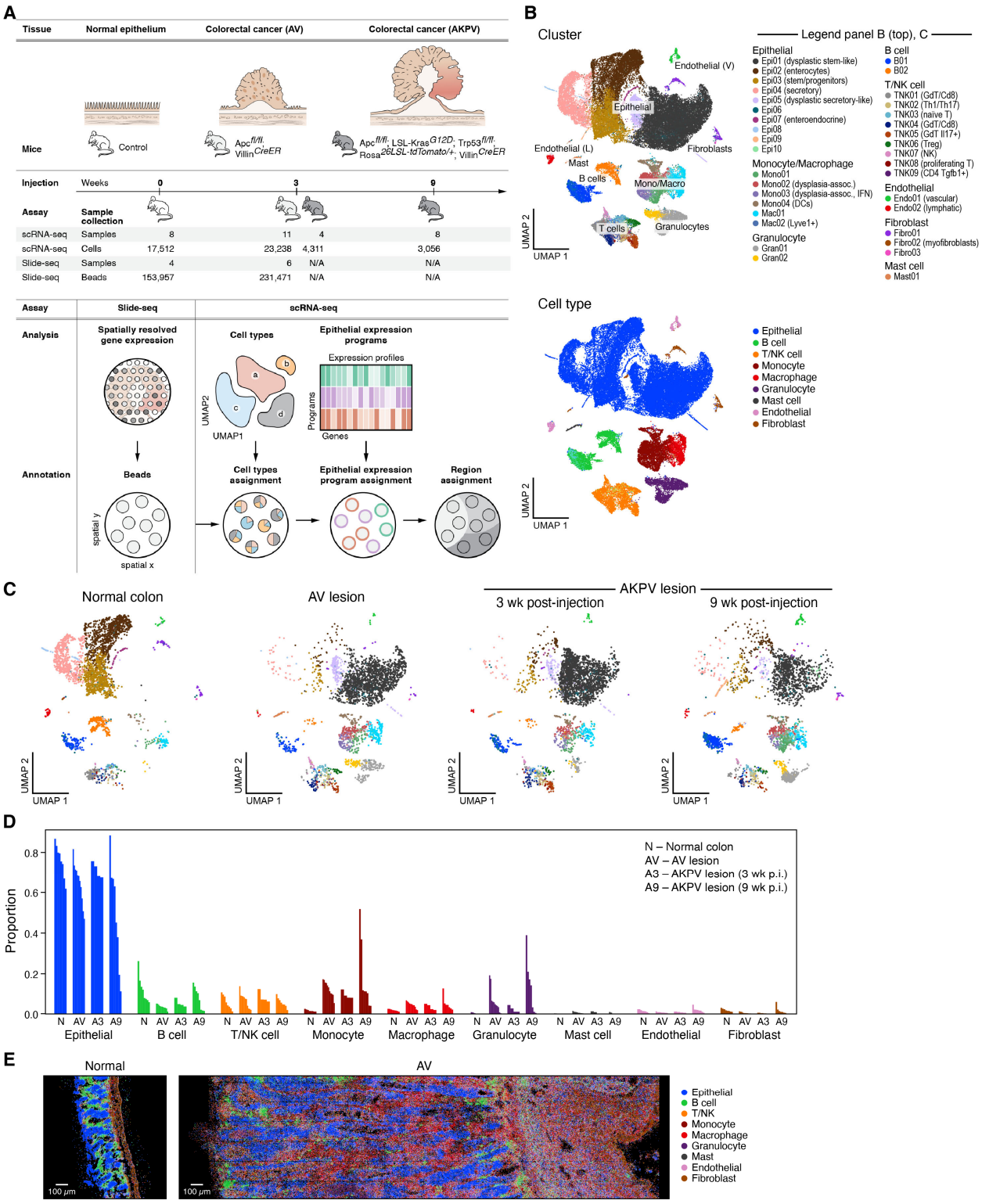
1515 94. Puram, S.V., Parikh, A.S., and Tirosh, I. (2018). Single cell RNA-seq highlights a role for a
1516 partial EMT in head and neck cancer. Mol Cell Oncol *5*, e1448244.

1517 95. Liu, J., Lichtenberg, T., Hoadley, K.A., Poisson, L.M., Lazar, A.J., Cherniack, A.D.,
1518 Kovatich, A.J., Benz, C.C., Levine, D.A., Lee, A.V., et al. (2018). An Integrated TCGA Pan-
1519 Cancer Clinical Data Resource to Drive High-Quality Survival Outcome Analytics. Cell *173*,
1520 400-416.e11.

1521 96. Davidson-Pilon, C. (2019). lifelines: survival analysis in Python. The Journal of Open Source
1522 Software *4*, 1317.

1523 97. Fazilaty, H., Brügger, M.D., Valenta, T., Szczerba, B.M., Berkova, L., Doumpas, N.,
1524 Hausmann, G., Scharl, M., and Basler, K. (2021). Tracing colonic embryonic transcriptional
1525 profiles and their reactivation upon intestinal damage. Cell Rep. *36*, 109484.

1526



1527

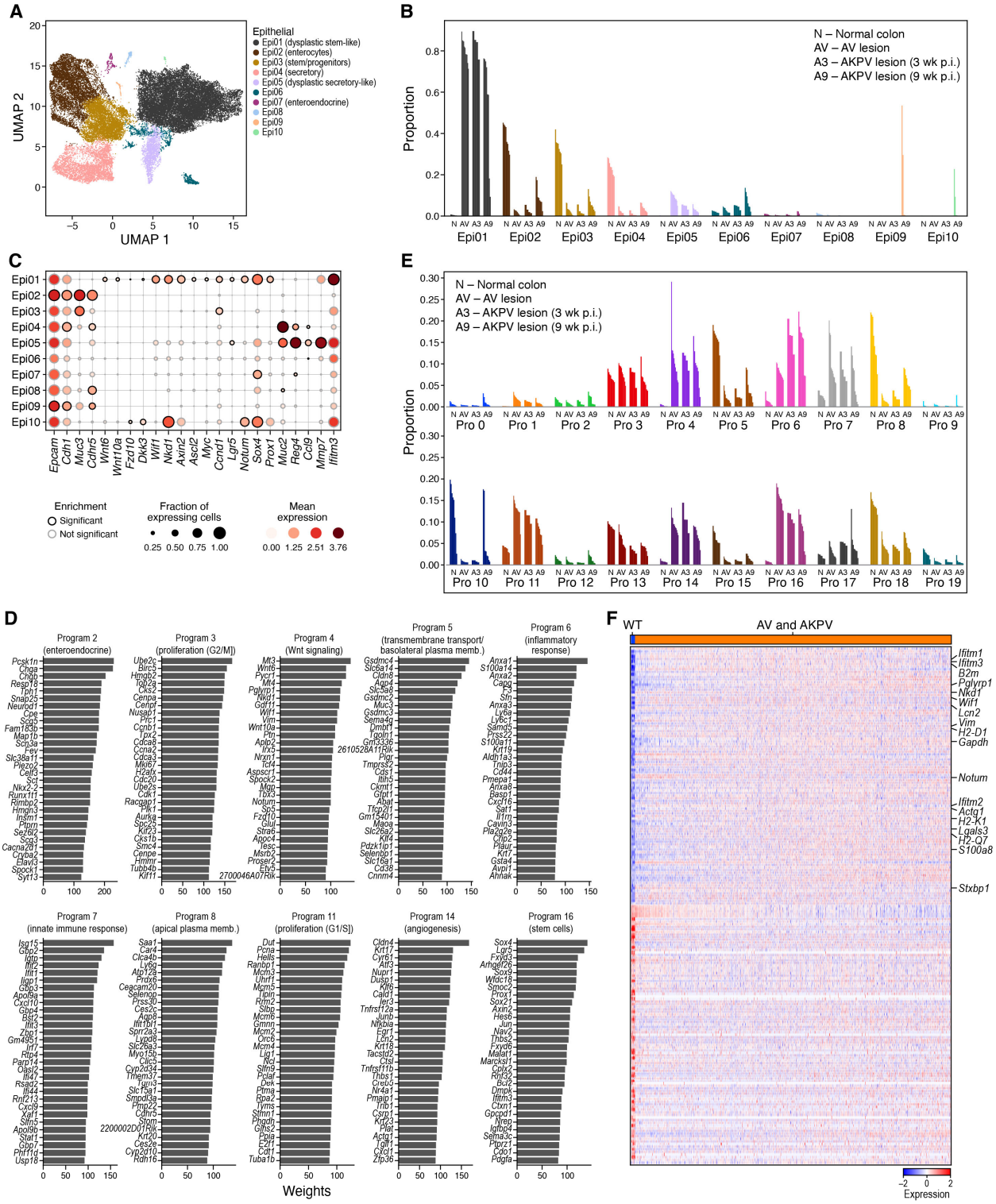
1528

1529

1530 **Figure 1. A single cell atlas of healthy colon and dysplastic lesions in mouse.**

1531 **A.** Study overview. **B.** Major cell subsets of healthy colon and dysplastic lesions. 2D embedding
1532 of 48,115 single cell profiles colored by cluster (top, legend) or annotated cell type (bottom,
1533 legend). **C,D.** Changes in cell composition in dysplastic tissues. C. 2D embedding of single cell
1534 profiles, showing only the cells in each condition state, subsampled to equal numbers of cells per
1535 condition state, colored by cluster (same legend as in B (top)). D. Proportion of cells (y axis) of
1536 each cell type in each sample (x axis). **E.** multiplex RNA *in situ* analysis. Representative images
1537 of Cartana analysis of normal colon (left) and AV lesions (right) colored by cell type assignment
1538 (same as **Supplementary item 2A**).

1539


A
UMAP 2
UMAP 1
Epithelial
Epi01 (dysplastic stem-like)
Epi02 (enterocytes)
Epi03 (stem/progenitors)
Epi04 (secretory)
Epi05 (dysplastic secretory-like)
Epi06
Epi07 (enteroendocrine)
Epi08
Epi09
Epi10
B
Proportion
N – Normal colon
AV – AV lesion
A3 – AKPV lesion (3 wk p.i.)
A9 – AKPV lesion (9 wk p.i.)
N AV A3 A9
Epi01 Epi02 Epi03 Epi04 Epi05 Epi06 Epi07 Epi08 Epi09 Epi10
C
Epi01 Epi02 Epi03 Epi04 Epi05 Epi06 Epi07 Epi08 Epi09 Epi10
Epcam Cdh1 Muc3 Cdhr5 Wnt6 Wnt10a Fzd10 Dkk3 Wif1 Nkd1 Axin2 Ascl2 Myc Ccnd1 Lgr5 Notum Sox4 Prox1 Muc2 Reg4 Ccl9 Mmp7 Ifitm3
Enrichment
Significant
Not significant
Fraction of expressing cells
0.25 0.50 0.75 1.00
Mean expression
0.00 1.25 2.51 3.76
E
Proportion
N – Normal colon
AV – AV lesion
A3 – AKPV lesion (3 wk p.i.)
A9 – AKPV lesion (9 wk p.i.)
N AV A3 A9
Pro 0 Pro 1 Pro 2 Pro 3 Pro 4 Pro 5 Pro 6 Pro 7 Pro 8 Pro 9
Pro 10 Pro 11 Pro 12 Pro 13 Pro 14 Pro 15 Pro 16 Pro 17 Pro 18 Pro 19
D
Program 2 (enteroendocrine)
Program 3 (proliferation (G2/M))
Program 4 (Wnt signaling)
Program 5 (transmembrane transport/ basolateral plasma memb.)
Program 6 (inflammatory response)
Program 7 (innate immune response)
Program 8 (apical plasma memb.)
Program 11 (proliferation (G1/S))
Program 14 (angiogenesis)
Program 16 (stem cells)
Weights
F
WT
AV and AKPV
Ifitm1
Ifitm3
B2m
Pglyrp1
Nkd1
Wif1
Lcn2
Vim
H2-D1
Gapdh
Notum
Ifitm2
Actg1
H2-K1
Lgals3
H2-Q7
S100a8
Stxbp1
Expression

1540

1541

1542 **Figure 2. Composition and cell intrinsic expression program changes in dysplastic epithelial**
1543 **cells.**

1544 **A-C.** Compositional changes in epithelial cells in dysplastic tissue. A. 2D embedding of epithelial
1545 cell profiles colored by clusters (legend). Cluster Epi06: potential doublets (**Methods**). B.
1546 Proportion of cells out of all epithelial cells (y axis) of each epithelial cell subset in each sample
1547 (x axis). C. Fraction of expressing cells (dot size) and mean expression in expressing cells (dot
1548 color) of marker genes (columns) for each cluster (rows). **D-E.** Use of epithelial cell programs
1549 changes in dysplastic tissue. D. Weights (x axis) of each of the 20 top ranked genes (y axis) for
1550 each program. E. Proportion of program weights summed over all epithelial cells (y axis) in each
1551 sample (x axis). **F.** Stem cell program 16 is induced in epithelial cells in dysplastic tissue. Scaled
1552 log-normalized expression (color bar) of the top 100 genes differentially expressed between cells
1553 from normal colon and from dysplastic (AV and AKPV) across the 10,812 cells that accounted for
1554 90% of program 16's expression across all epithelial cells (columns). Selected program genes are
1555 marked.

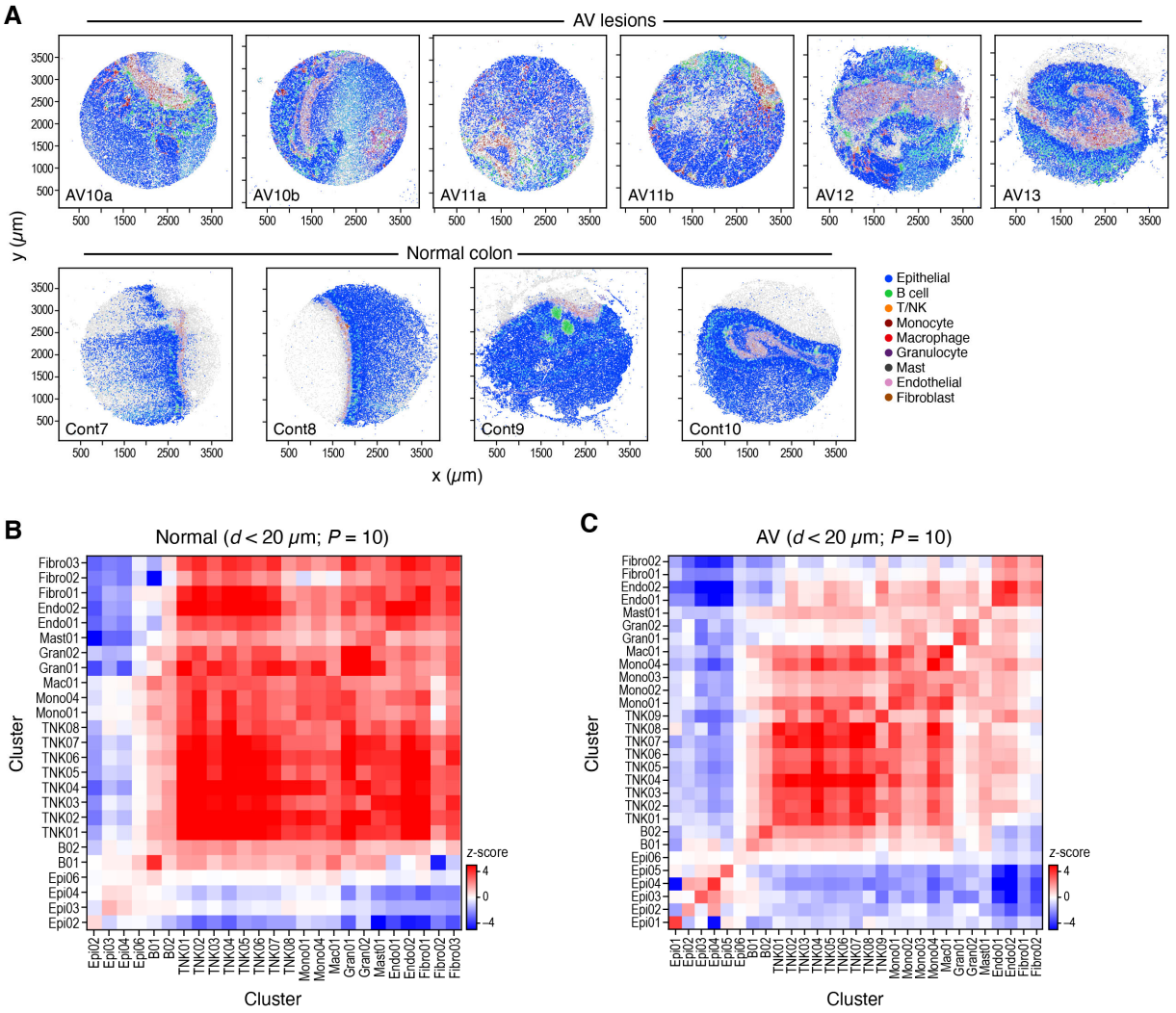
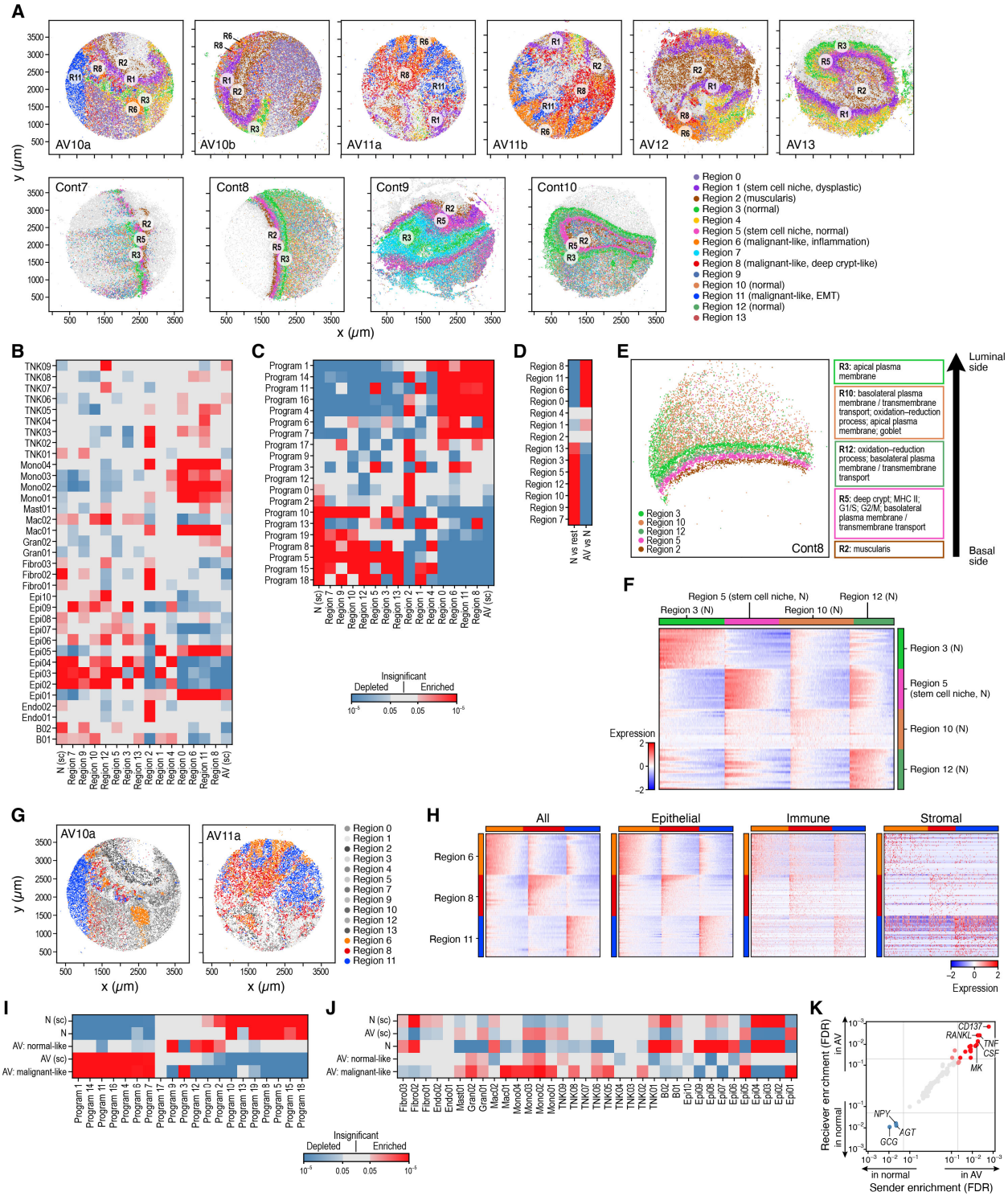


Figure 3. Altered cell type neighborship in CRC.

A. Cell type distributions *in situ*. Slide-seq pucks of dysplastic (top) and normal (bottom) tissue colored by TACCO assignment of cell labels (legend, light grey: low quality beads) (x and y axis: spatial coordinates in μm). **B,C.** Cell type neighborships in normal and dysplastic colon tissue. Short-range (up to 20μm) neighborship enrichment (Z score, color bar) vs. a background of spatially random annotation assignments for each pair of cell annotations (rows, columns) in normal (A) and dysplastic (B) tissue.


A
AV10a
AV10b
AV11a
AV11b
AV12
AV13
Cont7
Cont8
Cont9
Cont10
y (μm)
x (μm)
Region 0
Region 1 (stem cell niche, dysplastic)
Region 2 (muscularis)
Region 3 (normal)
Region 4
Region 5 (stem cell niche, normal)
Region 6 (malignant-like, inflammation)
Region 7
Region 8 (malignant-like, deep crypt-like)
Region 9
Region 10 (normal)
Region 11 (malignant-like, EMT)
Region 12 (normal)
Region 13
B
C
D
N vs rest
AV vs N
E
Luminal side
Basal side
R3: apical plasma membrane
R10: basolateral plasma membrane / transmembrane transport; oxidation–reduction process; apical plasma membrane; goblet
R12: oxidation–reduction process; basolateral plasma membrane / transmembrane transport
R5: deep crypt; MHC II; G1/S; G2/M; basolateral plasma membrane / transmembrane transport
R2: muscularis
F
Region 5 (stem cell niche, N)
Region 12 (N)
Region 3 (N)
Region 10 (N)
Expression
G
AV10a
AV11a
H
All
Epithelial
Immune
Stromal
Region 6
Region 8
Region 11
Expression
I
J
K
Insignificant
Depleted
Enriched
Reciever enrichment (FDR)
in AV
in normal
Sender enrichment (FDR)
CD137
RANKL
TNF
CSF
MK
NPY
AGT
GCG

1566 **Figure 4**. **Three cellular neighborhoods associated with tumor progression.**

1567 **A.** Spatial regions. Slide-seq pucks of AV (top) and normal (bottom) mouse colon colored by
1568 TACCO regions (legend, light gray: low quality beads) (x and y axis are spatial coordinates in
1569 μm). **B,C.** Enrichment and depletion of cell subsets and epithelial programs across different
1570 regions. Significance (FDR, color bar, two-sided Welch's t-test on CLR-transformed
1571 compositions) of enrichment (red) or depletion (blue) of specific cell subsets (rows, B) or epithelial
1572 cell programs (rows, C) in the different regions defined by TACCO (columns) as well as all normal
1573 ("N (ref.)", leftmost column) and AV ("AV (ref.)", leftmost column) samples. **D.** TACCO defined
1574 regions preferentially relate to normal or AV tissue. Significance (FDR, color bar, two-sided
1575 Welch's t-test on CLR-transformed compositions) of enrichment (red) or depletion (blue) of each
1576 TACCO defined region (rows) in normal ("N vs. rest") and AV ("AV vs. N") samples (columns).
1577 **E.** TACCO reveals normal colon architecture. Left: Slide-seq puck of normal mouse colon colored
1578 by TACCO region annotations (legend) (x and y axis: spatial coordinates (μm)). Right: Main
1579 epithelial expression programs enriched in each region (FDR<6.3 10^{-4}, two-sided Welch's t test
1580 on CLR-transformed compositions) except region 2 (muscularis), which is characterized by non-
1581 epithelial (stromal) cell types. **F.** Expression signatures of cells in normal regions 3,5,10 and 12.
1582 Scaled log-normalized expression of the top 20 differentially expressed genes (rows) for each bead
1583 (columns) in the region. **G,H**. Malignant-like regions. G. Slide-seq pucks of two AV lesions
1584 colored by TACCO annotations of malignant-like regions 6, 8 and 11. H. Scaled log-normalized
1585 expression of the top 20 differentially expressed genes (rows) of each bead (left, columns) in the
1586 region; or epithelial (middle left), immune (middle right) or stromal (right) fractions of beads
1587 (columns) in regions 6, 8 and 11 in dysplastic lesions. **I,J.** Epithelial cell subsets and programs
1588 associated with "malignant-like", "normal-like" and normal tissues. Significance (FDR, color bar,

1589 two-sided Welch's t-test on CLR-transformed compositions) of enrichment (red) or depletion
1590 (blue) of epithelial cell programs (I, rows) or epithelial, immune and stromal cell subsets (J, rows)
1591 in different tissue types (columns) based on Slide-seq or scRNA-seq ("sc") samples. **K.** Inferred
1592 interaction pathways. Enrichment (FDR) in AV *vs*. normal tissue of corresponding "sender" (x
1593 axis) and "receiver" (y axis) (aggregated over ligand-receptor pairs) pathways (dots). Red/blue:
1594 pathways significantly enriched in AV/normal samples and (light red: only for either sender or
1595 receiver). The top 5 enriched pathways (in each direction) are labeled.

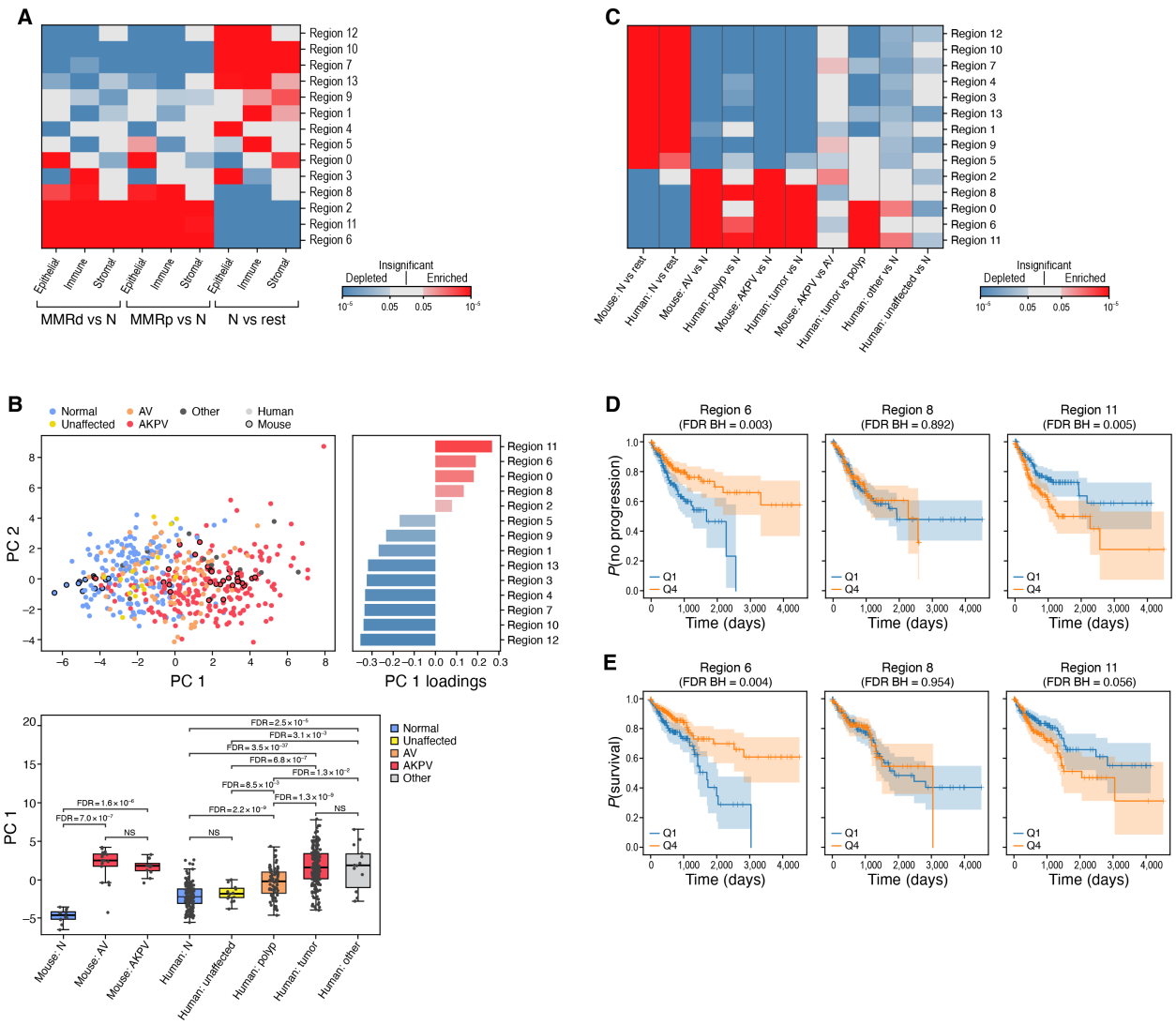
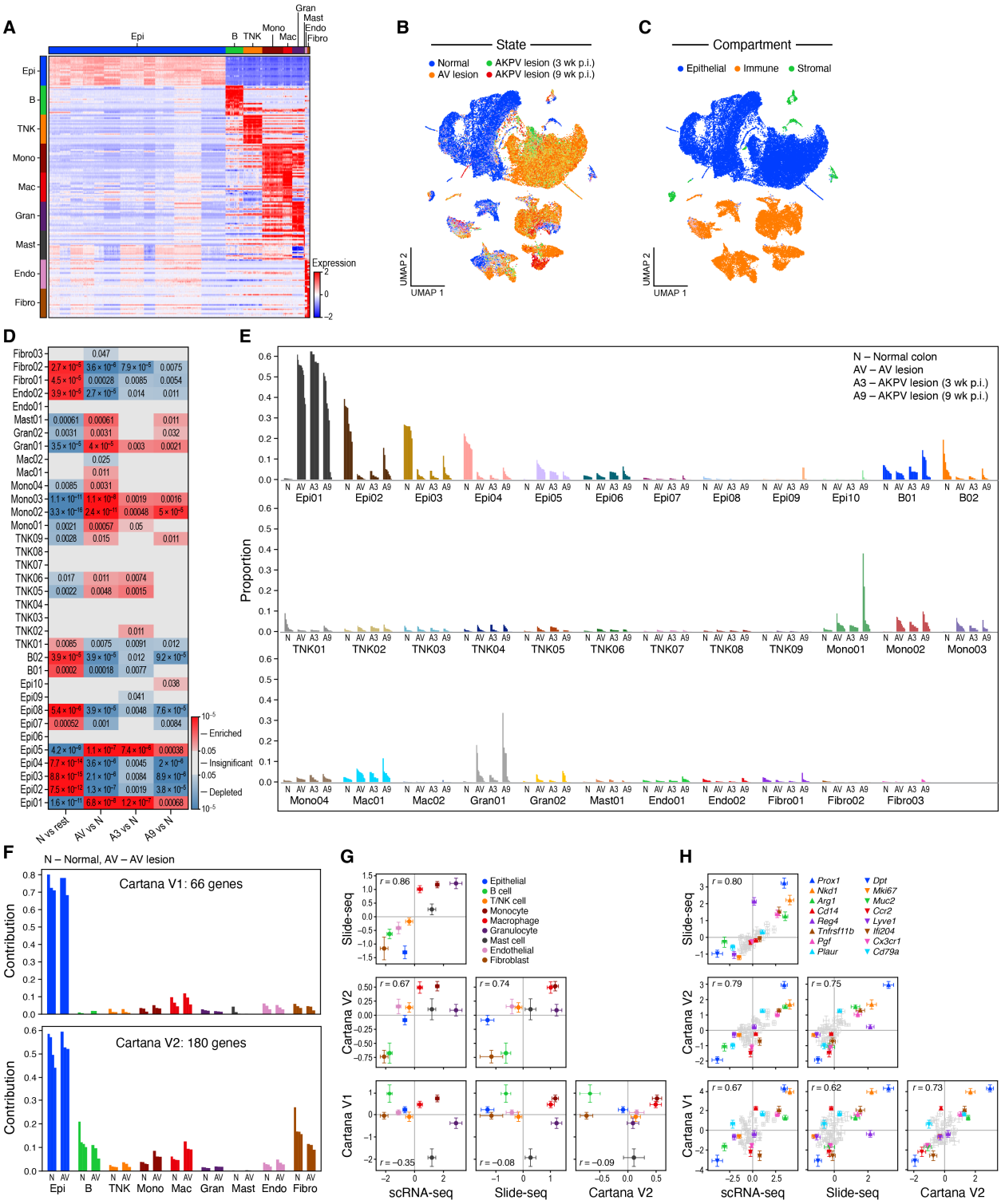


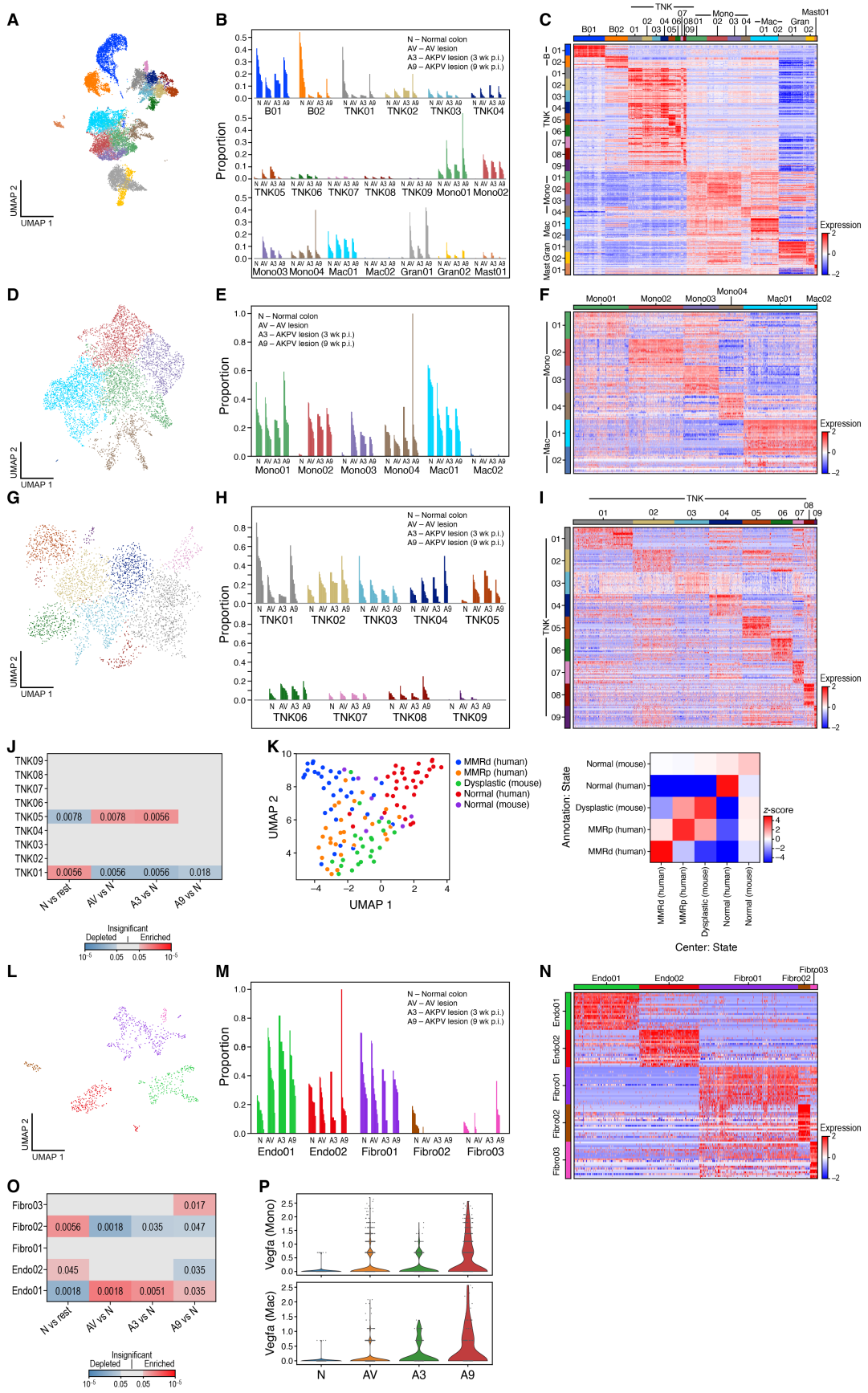
Figure 5. Mouse tumor regions associated with tumor progression in human colorectal tumors.

A. Expression profiles characterizing mouse regions are recapitulated in human tumors. Significance (FDR, color bar, two-sided Welch's t-test on CLR-transformed compositions) of enrichment (red) or depletion (blue) of region-associated epithelial, immune or stromal profiles (rows) compared between normal, MMRp, or MMRd samples (columns). **B,C.** Mouse regions capture malignant features in human tumors. B. Top left: First (PC1, x axis) and second (PC2, y axis) principal components of mouse region scores of mouse and human epithelial pseudo-bulk

1605 samples. Top right: PC1 loadings (x axis) of each mouse region score (y axis). Bottom: PC1 values
1606 (box plots show mean, quartiles, and whiskers for the full data distribution except for outliers
1607 outside 1.5 times the interquartile range (IQR)) for each type of mouse or human sample (x axis).
1608 C. Significance (FDR, color bar, two-sided Welch's t-test) of enrichment (red) or depletion (blue)
1609 of region-associated profile scores (rows) in normal and dysplastic samples (columns) in human
1610 or mouse. **D,E.** Expression of malignant like regions 6 and 11 in tumors are associated with PFI
1611 (D) and OS (E) in human patients. Kaplan-Meier PFI (e, n = 662[95]) or OS (f, n = 662[95]) analysis
1612 of human bulk RNA-seq cohort stratified by malignant-like region profile scores.


A
Epi
B
TNK
Mono
Mac
Gran
Mast
Endo
Fibro
Expression
B
State
Normal
AV lesion
AKPV lesion (3 wk p.i.)
AKPV lesion (9 wk p.i.)
UMAP 1
UMAP 2
C
Compartment
Epithelial
Immune
Stromal
D
N vs rest
AV vs N
A3 vs N
A9 vs N
Enriched
Insignificant
Depleted
E
N – Normal colon
AV – AV lesion
A3 – AKPV lesion (3 wk p.i.)
A9 – AKPV lesion (9 wk p.i.)
Proportion
F
N – Normal, AV – AV lesion
Cartana V1: 66 genes
Cartana V2: 180 genes
Contribution
G
Epithelial
B cell
T/NK cell
Monocyte
Macrophage
Granulocyte
Mast cell
Endothelial
Fibroblast
Slide-seq
Cartana V2
Cartana V1
scRNA-seq
H
Prox1
Nkd1
Arg1
Cd14
Reg4
Tnfrsf11b
Pgf
Plaur
Dpt
Mki67
Muc2
Ccr2
Lyve1
Ifi204
Cx3cr1
Cd79a

1616 **Figure S1**. **Marker genes, cell states and cell types in the healthy and dysplastic mouse**
1617 **colon atlas**.
1618 **A.** Cell-type expression signatures. Scaled log-normalized expression (color bar) of the top 20
1619 differentially expressed genes (rows) in cells (columns) from each cell type. **B,C.** Distinct
1620 condition states and compartments. 2D embedding of all single cell profiles (dots) colored by either
1621 condition state (B) or compartment (C). **D.** Changes in cell composition in dysplastic lesions.
1622 Significance (FDR, color bar, two-sided Welch's t-test on CLR-transformed compositions) of
1623 enrichment (red) or depletion (blue) of each cell subset (rows) between samples from different
1624 conditions (columns). **E.** Changes in cell composition between healthy and dysplastic tissue.
1625 Proportion of cells (y axis) from each cell subset (x axis) out of all cells in each sample (x axis).
1626 **F-H.** Multiplex *in situ* RNA profiles (Cartana) reproduce scRNA-seq and Slide-Seq findings. F.
1627 Proportion of RNA molecules (y axis) attributed to each cell type in each sample (x axis) in two
1628 Cartana experiments, with different gene panels. G. Change in cell proportions in AV vs. normal
1629 samples (difference of CLR-transformed cell type fractions, x and y axes) for each cell type (dots,
1630 color) based on scRNA-seq, Slide-Seq or Cartana (axis labels). Error bars: bootstrapped standard
1631 error of the mean (normal n=8, 16, 12, 16 and AV n=11, 24, 12, 12 for scRNA-seq, Slide-seq,
1632 Cartana V1, Cartana V2). H. Change in marker gene expression (dots) in AV vs. normal samples
1633 (difference of CLR-transformed gene fractions, x and y axes) based on scRNA-seq, Slide-Seq or
1634 Cartana (axis labels), for genes measured by all 3 methods. Genes with the 8 maximal and minimal
1635 expression ranks across methods are labeled. Error bars: bootstrapped standard error of the mean
1636 (normal n=8, 16, 12, 16 and AV n=11, 24, 12, 12 for scRNA-seq, Slide-seq, Cartana V1, Cartana
1637 V2).

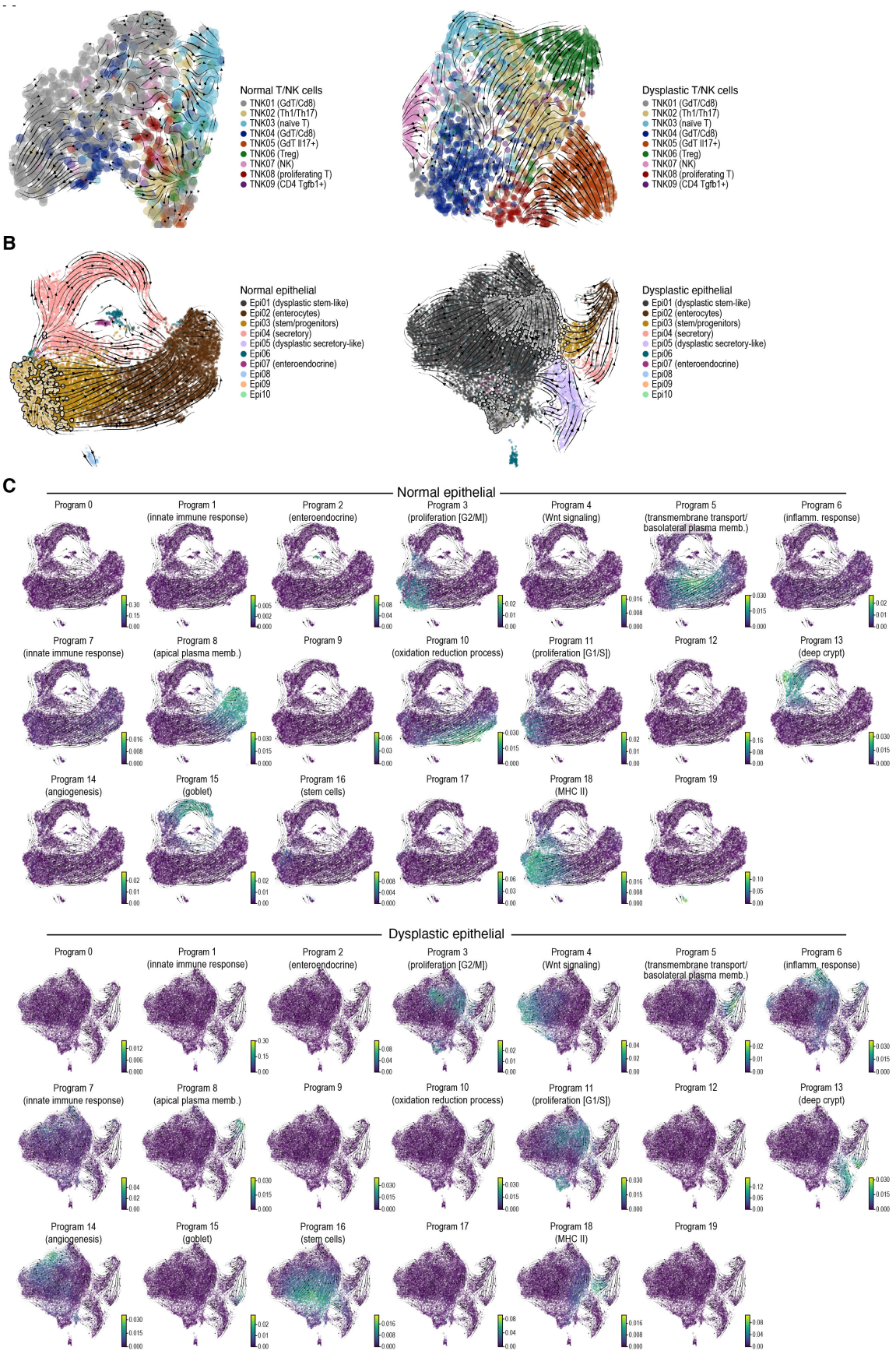

A
B
C
D
E
F
G
H
I
J
K
L
M
N
O
P
UMAP 1
UMAP 2
Proportion
N – Normal colon
AV – AV lesion
A3 – AKPV lesion (3 wk p.i.)
A9 – AKPV lesion (9 wk p.i.)
B01 B02 TNK01 TNK02 TNK03 TNK04
TNK05 TNK06 TNK07 TNK08 TNK09 Mono01 Mono02
Mono03 Mono04 Mac01 Mac02 Gran01 Gran02 Mast01
Expression
Mono01 Mono02 Mono03 Mono04 Mac01 Mac02
TNK01 TNK02 TNK03 TNK04 TNK05
TNK06 TNK07 TNK08 TNK09
N vs rest
AV vs N
A3 vs N
A9 vs N
Insignificant
Depleted
Enriched
MMRd (human)
MMRp (human)
Dysplastic (mouse)
Normal (human)
Normal (mouse)
Annotation: State
Center: State
z-score
Endo01 Endo02 Fibro01 Fibro02 Fibro03
Vegfa (Mono)
Vegfa (Mac)
N AV A3 A9

1638

1639 **Figure S2. Compositional and cell intrinsic changes in stromal and immune cells.**

1640 **A-C.** Immune cell subsets and composition. A. 2D embedding of immune cell profiles colored by
1641 clusters (as in B). B. Proportion of cells out of all immune cells (y axis) of each immune cell subset
1642 in each sample (x axis). C. Scaled log-normalized expression (color bar) of the top 20 differentially
1643 expressed genes (rows) in cells (columns) from each immune cell subset (color bar on top). **D-F**.
1644 Monocyte and macrophage cell subsets and composition. D. 2D embedding of monocyte and
1645 macrophage cell profiles colored by clusters (as in E). E. Proportion of cells out of all monocytes
1646 and macrophages (y axis) of each monocyte and macrophage cell subset in each sample (x axis).
1647 F. Scaled log-normalized expression (color bar) of the top 20 differentially expressed genes (rows)
1648 in cells (columns) from each monocyte and macrophage cell subset (color bar on top). **G-I.** T/NK
1649 cell subsets and composition. G. 2D embedding of T/NK cell profiles colored by clusters (as in
1650 H). H. Proportion of cells out of all T/NK cells (y axis) of each T/NK cell subset in each sample
1651 (x axis). I. Scaled log-normalized expression (color bar) of the top 20 differentially expressed
1652 genes (rows) in cells (columns) from each T/NK cell subset (color bar on top). **J.** Enrichment of
1653 IL17+ γδT cells (TNK05) and depletion of CD8+ γδT cells (TNK01) in dysplastic lesions.
1654 Significance (FDR, color bar, two-sided Welch's t-test on CLR-transformed compositions) of
1655 enrichment (red) or depletion (blue) of each T/NK cell subset (rows) between samples from
1656 different conditions (columns). **K**. T/NK cell compositions is similar in human and mouse. Left:
1657 2D embedding of T/NK cell composition profiles of human and mouse samples colored by sample
1658 type (legend) (**STAR Methods**). Right: Similarity of T/NK cell composition (enrichment z-scores)
1659 in the 2D embedding between each set of samples (rows, columns). **L-N.** Stromal cell subsets and
1660 composition. L. 2D embedding of stromal cell profiles colored by clusters (legend). M. Proportion
1661 of cells out of all stromal cells (y axis) of each stromal cell subset in each sample (x axis). N.

1662 Scaled log-normalized expression (color bar) of the top 20 differentially expressed genes (rows)
1663 in cells (columns) from each stromal cell subset (color bar on top). **O**. Enrichment of vascular
1664 endothelial cells (Endo01) and depletion of myofibroblasts (Fibro02) in dysplastic lesions.
1665 Significance (FDR, color bar, two-sided Welch's t test on CLR-transformed compositions) of
1666 enrichment (red) or depletion (blue) of each stromal cell subset (rows) between samples from
1667 different conditions (columns). **P.** Increased VEGFA expression in monocyte-macrophage
1668 populations with dysplasia. Distribution of expression (y axis, log1p(counts)) of *VegfA* in
1669 monocytes and macrophages from different conditions (x axis).

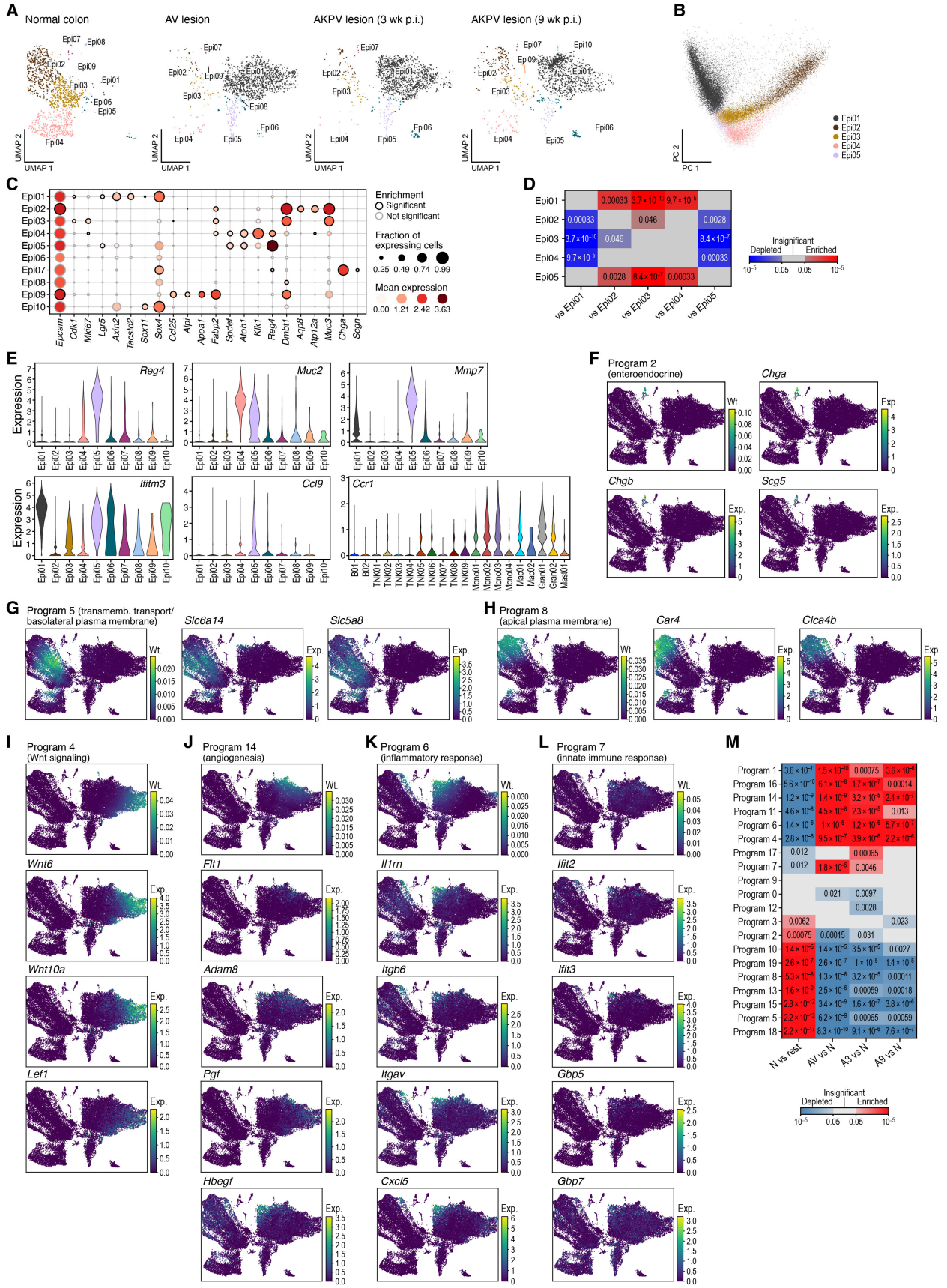

Normal T/NK cells
TNK01 (GdT/Cd8)
TNK02 (Th1/Th17)
TNK03 (naïve T)
TNK04 (GdT/Cd8)
TNK05 (GdT Il17+)
TNK06 (Treg)
TNK07 (NK)
TNK08 (proliferating T)
TNK09 (CD4 Tgfb1+)
Dysplastic T/NK cells
TNK01 (GdT/Cd8)
TNK02 (Th1/Th17)
TNK03 (naïve T)
TNK04 (GdT/Cd8)
TNK05 (GdT Il17+)
TNK06 (Treg)
TNK07 (NK)
TNK08 (proliferating T)
TNK09 (CD4 Tgfb1+)
B
Normal epithelial
Epi01 (dysplastic stem-like)
Epi02 (enterocytes)
Epi03 (stem/progenitors)
Epi04 (secretory)
Epi05 (dysplastic secretory-like)
Epi06
Epi07 (enteroendocrine)
Epi08
Epi09
Epi10
Dysplastic epithelial
Epi01 (dysplastic stem-like)
Epi02 (enterocytes)
Epi03 (stem/progenitors)
Epi04 (secretory)
Epi05 (dysplastic secretory-like)
Epi06
Epi07 (enteroendocrine)
Epi08
Epi09
Epi10
C
Normal epithelial
Program 0
Program 1 (innate immune response)
Program 2 (enteroendocrine)
Program 3 (proliferation [G2/M])
Program 4 (Wnt signaling)
Program 5 (transmembrane transport/ basolateral plasma memb.)
Program 6 (inflamm. response)
Program 7 (innate immune response)
Program 8 (apical plasma memb.)
Program 9
Program 10 (oxidation reduction process)
Program 11 (proliferation [G1/S])
Program 12
Program 13 (deep crypt)
Program 14 (angiogenesis)
Program 15 (goblet)
Program 16 (stem cells)
Program 17
Program 18 (MHC II)
Program 19
Dysplastic epithelial
Program 0
Program 1 (innate immune response)
Program 2 (enteroendocrine)
Program 3 (proliferation [G2/M])
Program 4 (Wnt signaling)
Program 5 (transmembrane transport/ basolateral plasma memb.)
Program 6 (inflamm. response)
Program 7 (innate immune response)
Program 8 (apical plasma memb.)
Program 9
Program 10 (oxidation reduction process)
Program 11 (proliferation [G1/S])
Program 12
Program 13 (deep crypt)
Program 14 (angiogenesis)
Program 15 (goblet)
Program 16 (stem cells)
Program 17
Program 18 (MHC II)
Program 19

1670

1671 **Figure S3. Changes in T/NK and epithelial cell differentiation in dysplastic lesions.**

1672 **A.** T/NK cell differentiation. 2D embedding of T/NK cell profiles from normal tissues (left) and
1673 dysplastic lesions (right), colored by cell subset. Streamlines: averaged and projected RNA
1674 velocities. **B,C.** Proliferating stem-like cells give rise to an expanded stem-like compartment and
1675 differentiated-like tumor cells in dysplastic lesions. 2D embedding of epithelial cell profiles from
1676 normal tissues (B, left and C, top) and dysplastic lesions (B, right, and C, bottom), colored by cell
1677 subset (B), or by expression of epithelial programs (C). Streamlines: averaged and projected RNA
1678 velocities. Outlined dots (B): Proliferative cells (cells with more than 50% program weight in the
1679 proliferation programs #3 and #11).

1680

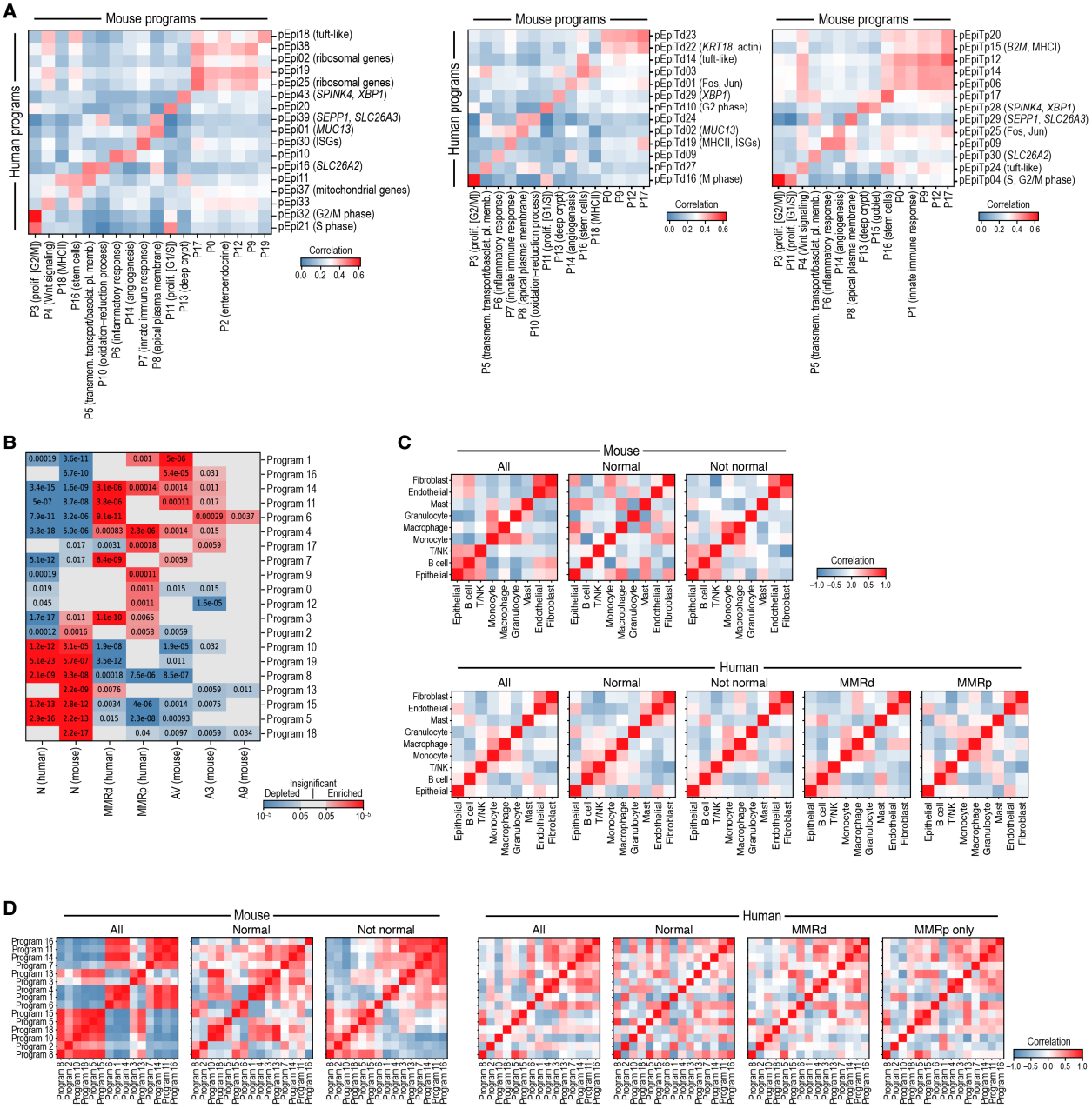

A
Normal colon
AV lesion
AKPV lesion (3 wk p.i.)
AKPV lesion (9 wk p.i.)
UMAP 1
UMAP 2
B
PC 1
PC 2
Epi01
Epi02
Epi03
Epi04
Epi05
C
Enrichment
Significant
Not significant
Fraction of expressing cells
0.25 0.49 0.74 0.99
Mean expression
0.00 1.21 2.42 3.63
Epcam
Cdk1
Mki67
Lgr5
Axin2
Tacstd2
Sox11
Sox4
Ccl25
Alpi
Apoa1
Fabp2
Spdef
Atoh1
Klk1
Reg4
Dmbt1
Aqp8
Atp12a
Muc3
Chga
Scgn
D
vs Epi01
vs Epi02
vs Epi03
vs Epi04
vs Epi05
Insignificant
Depleted
Enriched
E
Expression
Reg4
Muc2
Mmp7
Ifitm3
Ccl9
Ccr1
F
Program 2
(enteroendocrine)
Chga
Chgb
Scg5
Wt.
Exp.
G
Program 5 (transmemb. transport/
basolateral plasma membrane)
Slc6a14
Slc5a8
H
Program 8
(apical plasma membrane)
Car4
Clca4b
I
Program 4
(Wnt signaling)
Wnt6
Wnt10a
Lef1
J
Program 14
(angiogenesis)
Flt1
Adam8
Pgf
Hbegf
K
Program 6
(inflammatory response)
Il1rn
Itgb6
Itgav
Cxcl5
L
Program 7
(innate immune response)
Ifit2
Ifit3
Gbp5
Gbp7
M
Program 1
Program 16
Program 14
Program 11
Program 6
Program 4
Program 17
Program 7
Program 9
Program 0
Program 12
Program 3
Program 2
Program 10
Program 19
Program 8
Program 13
Program 15
Program 5
Program 18
N vs rest
AV vs N
A3 vs N
A9 vs N
Insignificant
Depleted
Enriched

1681

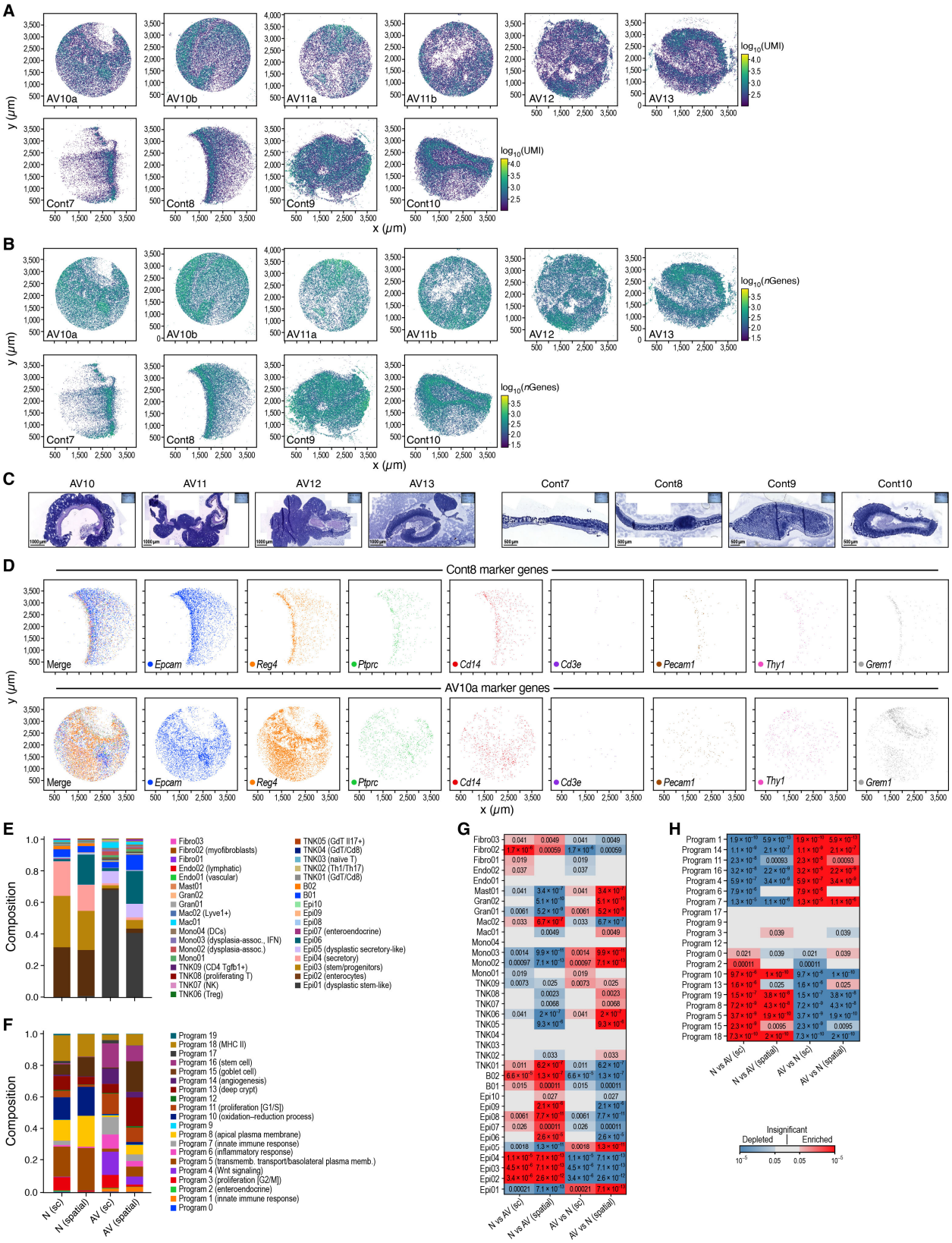
1682 **Figure S4. Changes in cell composition and expression programs usage in dysplastic**
1683 **epithelium.**

1684 **A-D.** Changes in epithelial cell composition in dysplastic tissues. A. 2D embedding of all single
1685 cell epithelial profiles showing only the profiles (dots) of cells from each condition state,
1686 subsampled to equal numbers of cells per condition state, colored by cluster. B. PCA captures
1687 differentiation status and lineage of epithelial clusters. First (PC1, x axis) and second (PC2, y axis)
1688 principal components of the expression of epithelial clusters. C. Mean expression (dot color) and
1689 fraction of expressing cells (dot size) of epithelial cell marker genes[97] (columns) in each cluster
1690 (rows). D. Significance (FDR, color bar, two-sided Welch's t test on ALR-transformed
1691 compositions with all non-tdTomato counts used as reference compartment) of enrichment (red)
1692 or depletion (blue) of tdTomato expression in cells from AKPVT samples between every pair of
1693 epithelial clusters. **E.** Dysplastic secretory like (Epi05) and immune cells express tumor-related
1694 genes. Distribution of expression (y axis, log1p(counts)) of different marker genes in cells from
1695 each epithelial/immune cell cluster (x axis). **F-L**. Epithelial gene programs. 2D embedding of all
1696 epithelial cells colored by the weight of each program (color bar) and the expression of selected
1697 program genes (color bar). **M.** Epithelial program characteristics of normal and dysplastic colon.
1698 Heatmap significance (FDR, color bar, two-sided Welch's t test on CLR-transformed
1699 compositions) of enrichment (red) or depletion (blue) of each epithelial program (rows) between
1700 normal vs. dysplastic tissues (columns).

1701

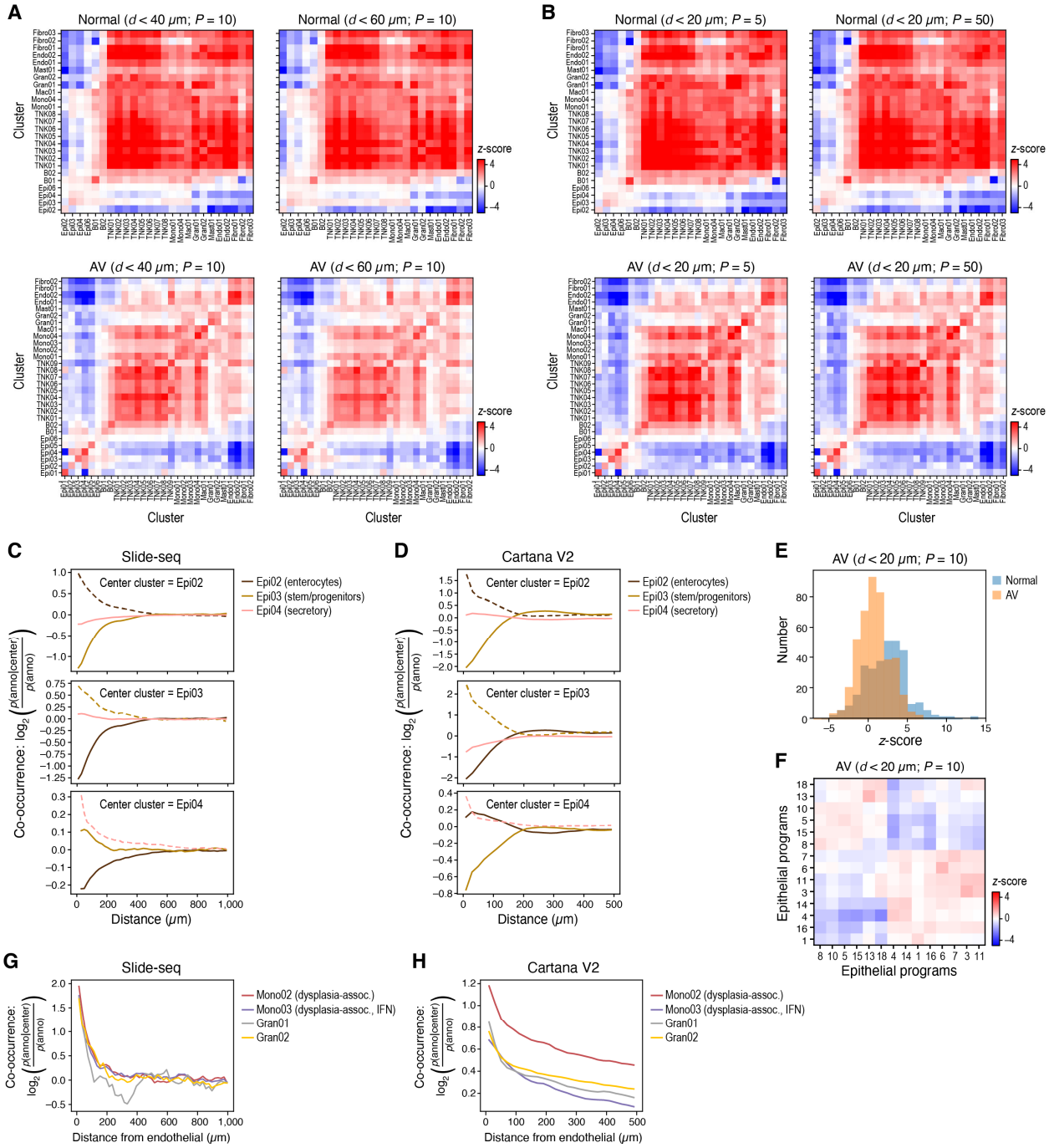

A
Mouse programs
Human programs
pEpi18 (tuft-like)
pEpi38
pEpi02 (ribosomal genes)
pEpi19
pEpi25 (ribosomal genes)
pEpi43 (SPINK4, XBP1)
pEpi20
pEpi39 (SEPP1, SLC26A3)
pEpi01 (MUC13)
pEpi30 (ISGs)
pEpi10
pEpi16 (SLC26A2)
pEpi11
pEpi37 (mitochondrial genes)
pEpi33
pEpi32 (G2/M phase)
pEpi21 (S phase)
P3 (prolif. [G2/M])
P4 (Wnt signaling)
P18 (MHCII)
P16 (stem cells)
P5 (transmem. transport/basolat. pl. memb.)
P10 (oxidation–reduction process)
P6 (inflammatory response)
P14 (angiogenesis)
P7 (innate immune response)
P8 (apical plasma membrane)
P11 (prolif. [G1/S])
P13 (deep crypt)
P17
P0
P2 (enteroendocrine)
P12
P9
P19
Correlation
0.0 0.2 0.4 0.6
pEpiTd23
pEpiTd22 (KRT18, actin)
pEpiTd14 (tuft-like)
pEpiTd03
pEpiTd01 (Fos, Jun)
pEpiTd29 (XBP1)
pEpiTd10 (G2 phase)
pEpiTd24
pEpiTd02 (MUC13)
pEpiTd19 (MHCII, ISGs)
pEpiTd09
pEpiTd27
pEpiTd16 (M phase)
P3 (prolif. [G2/M])
P5 (transmem. transport/basolat. pl. memb.)
P6 (inflammatory response)
P7 (innate immune response)
P8 (apical plasma membrane)
P10 (oxidation–reduction process)
P11 (prolif. [G1/S])
P13 (deep crypt)
P14 (angiogenesis)
P16 (stem cells)
P18 (MHCII)
P0
P9
P12
P17
pEpiTp20
pEpiTp15 (B2M, MHCI)
pEpiTp12
pEpiTp14
pEpiTp06
pEpiTp17
pEpiTp28 (SPINK4, XBP1)
pEpiTp29 (SEPP1, SLC26A3)
pEpiTp25 (Fos, Jun)
pEpiTp09
pEpiTp30 (SLC26A2)
pEpiTp24 (tuft-like)
pEpiTp04 (S, G2/M phase)
P3 (prolif. [G2/M])
P11 (prolif. [G1/S])
P4 (Wnt signaling)
P5 (transmem. transport/basolat. pl. memb.)
P6 (inflammatory response)
P14 (angiogenesis)
P8 (apical plasma membrane)
P13 (deep crypt)
P15 (goblet)
P16 (stem cells)
P0
P1 (innate immune response)
P9
P12
P17
B
Program 1
Program 16
Program 14
Program 11
Program 6
Program 4
Program 17
Program 7
Program 9
Program 0
Program 12
Program 3
Program 2
Program 10
Program 19
Program 8
Program 13
Program 15
Program 5
Program 18
N (human)
N (mouse)
MMRd (human)
MMRp (human)
AV (mouse)
A3 (mouse)
A9 (mouse)
Insignificant
Depleted
Enriched
C
Mouse
All
Normal
Not normal
Fibroblast
Endothelial
Mast
Granulocyte
Macrophage
Monocyte
T/NK
B cell
Epithelial
Correlation
−1.0 −0.5 0.0 0.5 1.0
Human
All
Normal
Not normal
MMRd
MMRp
D
Mouse
All
Normal
Not normal
Human
All
Normal
MMRd
MMRp only
Correlation
−1.0 −0.5 0.0 0.5 1.0

1708 **Figure S5. Conservation of cellular composition and expression programs between mouse**
1709 **and human scRNA-seq data.**
1710 **A,B.** Conservation of general and subtype-specific human and mouse epithelial expression
1711 programs. A. Pearson correlation coefficients (color) between program-specific expression
1712 profiles (**Methods**) of human (all samples, left, MMRd, middle and MMRp, right)) (rows) and
1713 mouse programs (columns). B. Mouse epithelial program enrichments in mouse and human tumor
1714 and normal samples. Significance (FDR, color bar, two-sided Welch t-test on CLR-transformed
1715 compositions) of enrichment (red) or depletion (blue) of mouse epithelial program expression
1716 (rows) in sample classes from human or mouse (columns). **C,D.** Human-mouse conservation of
1717 cell type and program associations. Pearson correlation coefficients (color) of the CLR-
1718 transformed cell type (C) or epithelial program (D) compositions across samples in mouse (left)
1719 or human (right) single cell data. In (D), data are hierarchically clustered for the "not normal"
1720 mouse case (AV and AKPV) and this ordering is applied to all other panels.


A
AV10a
AV10b
AV11a
AV11b
AV12
AV13
Cont7
Cont8
Cont9
Cont10
y (μm)
x (μm)
log10(UMI)
B
log10(nGenes)
C
AV10
AV11
AV12
AV13
Cont7
Cont8
Cont9
Cont10
D
Cont8 marker genes
AV10a marker genes
Merge
Epcam
Reg4
Ptprc
Cd14
Cd3e
Pecam1
Thy1
Grem1
E
Composition
Fibro03
Fibro02 (myofibroblasts)
Fibro01
Endo02 (lymphatic)
Endo01 (vascular)
Mast01
Gran02
Gran01
Mac02 (Lyve1+)
Mac01
Mono04 (DCs)
Mono03 (dysplasia-assoc., IFN)
Mono02 (dysplasia-assoc.)
Mono01
TNK09 (CD4 Tgfb1+)
TNK08 (proliferating T)
TNK07 (NK)
TNK06 (Treg)
TNK05 (GdT Il17+)
TNK04 (GdT/Cd8)
TNK03 (naive T)
TNK02 (Th1/Th17)
TNK01 (GdT/Cd8)
B02
B01
Epi10
Epi09
Epi08
Epi07 (enteroendocrine)
Epi06
Epi05 (dysplastic secretory-like)
Epi04 (secretory)
Epi03 (stem/progenitors)
Epi02 (enterocytes)
Epi01 (dysplastic stem-like)
F
Program 19
Program 18 (MHC II)
Program 17
Program 16 (stem cell)
Program 15 (goblet cell)
Program 14 (angiogenesis)
Program 13 (deep crypt)
Program 12
Program 11 (proliferation [G1/S])
Program 10 (oxidation–reduction process)
Program 9
Program 8 (apical plasma membrane)
Program 7 (innate immune response)
Program 6 (inflammatory response)
Program 5 (transmemb. transport/basolateral plasma memb.)
Program 4 (Wnt signaling)
Program 3 (proliferation [G2/M])
Program 2 (enteroendocrine)
Program 1 (innate immune response)
Program 0
N (sc)
N (spatial)
AV (sc)
AV (spatial)
G
N vs AV (sc)
N vs AV (spatial)
AV vs N (sc)
AV vs N (spatial)
H
Insignificant
Depleted
Enriched

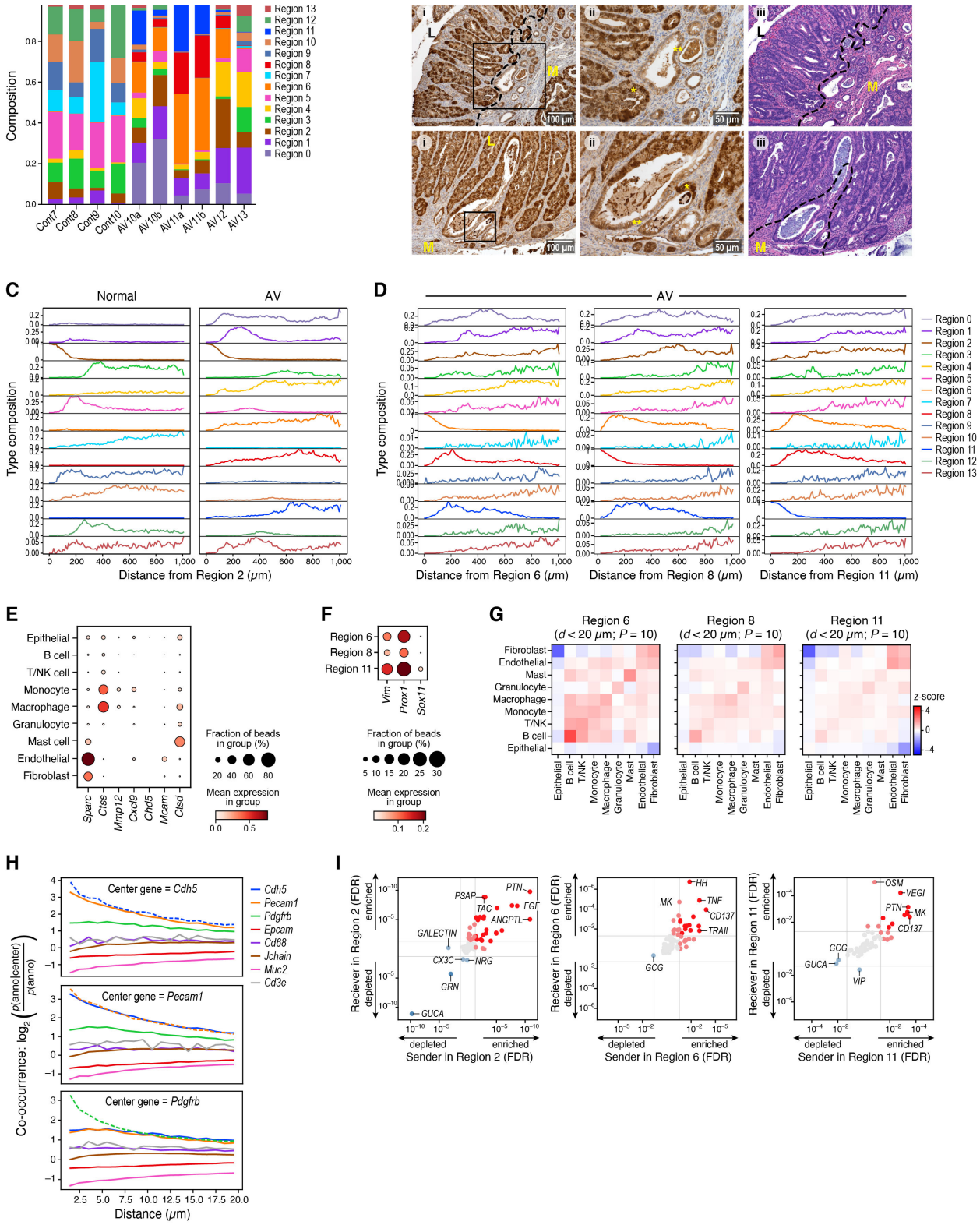
1723 **Figure S6**. **Spatial distributions of cells and programs across regions.**

1724 **A-C.** Slide-seq quality controls. Slide-seq pucks of AV (top) and normal (bottom) mouse colon
1725 colored by number of UMIs (A) or of genes (B) per bead. (x and y axis: spatial coordinates (μm)).
1726 C. H&E staining of sections adjacent to those used for Slide-seq. **D.** Selected marker gene
1727 expression. Slide-seq pucks from a normal (top) and AV (bottom) sample, colored by marker gene
1728 detected per bead. **E,F.** Spatial mapping of cell types and programs yields comparable composition
1729 to scRNA-seq. Distribution of the proportion (y axis) of contributions to each cell type (D) or
1730 program (E; based on fractional annotations) in cells (for scRNA-seq; "sc") or beads (for Slide-
1731 seq; "spatial"; based on fractional annotations) in samples from normal (N) or AV (AV) tissue.
1732 **G,H.** Distinct cell types and programs associated with AV and normal colon. Significance (FDR,
1733 color bar, two-sided Welch's t test on CLR-transformed compositions) of enrichment (red) or
1734 depletion (blue) of cell types (F, rows) or epithelial programs (G, rows) in normal (N) or AV (AV)
1735 tissues based on Slide-seq ("spatial") data or scRNA-seq ("sc").


A
Normal (d < 40 μm; P = 10)
Normal (d < 60 μm; P = 10)
AV (d < 40 μm; P = 10)
AV (d < 60 μm; P = 10)
B
Normal (d < 20 μm; P = 5)
Normal (d < 20 μm; P = 50)
AV (d < 20 μm; P = 5)
AV (d < 20 μm; P = 50)
Cluster
z-score
C
Slide-seq
Center cluster = Epi02
Center cluster = Epi03
Center cluster = Epi04
Epi02 (enterocytes)
Epi03 (stem/progenitors)
Epi04 (secretory)
Co-occurrence: log2 (p(anno|center) / p(anno))
Distance (μm)
D
Cartana V2
E
AV (d < 20 μm; P = 10)
Normal
AV
Number
z-score
F
Epithelial programs
G
Slide-seq
Mono02 (dysplasia-assoc.)
Mono03 (dysplasia-assoc., IFN)
Gran01
Gran02
Co-occurrence: log2 (p(anno|center) / p(anno))
Distance from endothelial (μm)
H
Cartana V2

1741 **Figure S7. Distinct cellular layout in normal and AV tissues**.

1742 **A,B.** Neighborship analysis is robust to the distance and number of randomizations. Neighborship
1743 enrichment (z-scores, color) *vs*. a background of spatially random annotation assignments for each
1744 pair of cell annotations (rows, columns) in normal (top) and AV (bottom) samples at varying
1745 distance (a, left: ≤40μm; right: ≤60 μm) or number of permutations (B, left: 5, right: 50). **C,D.**
1746 Reproducible spatial arrangement of epithelial cells in normal colon in Slide-seq and Cartana V2.
1747 Co-occurrence (y axis) of normal epithelial cell types (color) at different distances (x axis) from
1748 different central normal epithelial cell type (annotated on top) in Slide-Seq (C) and Cartana (D)
1749 data. **E.** Decreased spatial order of cell types in AV vs. normal tissue. Distribution of enrichment
1750 z-values (x axis) of cluster-cluster interactions (as shown in **Figure 3B,C**) in AV lesions (orange)
1751 and normal colon (blue). **F.** Epithelial program neighborships in AV tissue. Short-range (≤20μm)
1752 neighborship enrichment z-scores (color) *vs*. a background of spatially random annotation
1753 assignments for each pair of epithelial program annotations (rows, columns) in AV lesions. **G,H.**
1754 Reproducible spatial arrangement of dysplasia-associated immune cells relative to endothelial
1755 cells in Slide-seq and Cartana V2. Co-occurrence (y axis) of dysplasia-associated monocyte or
1756 granulocyte cells (color) with endothelial cells at different distances (x axis) in Slide-Seq (G) and
1757 Cartana (H) data.

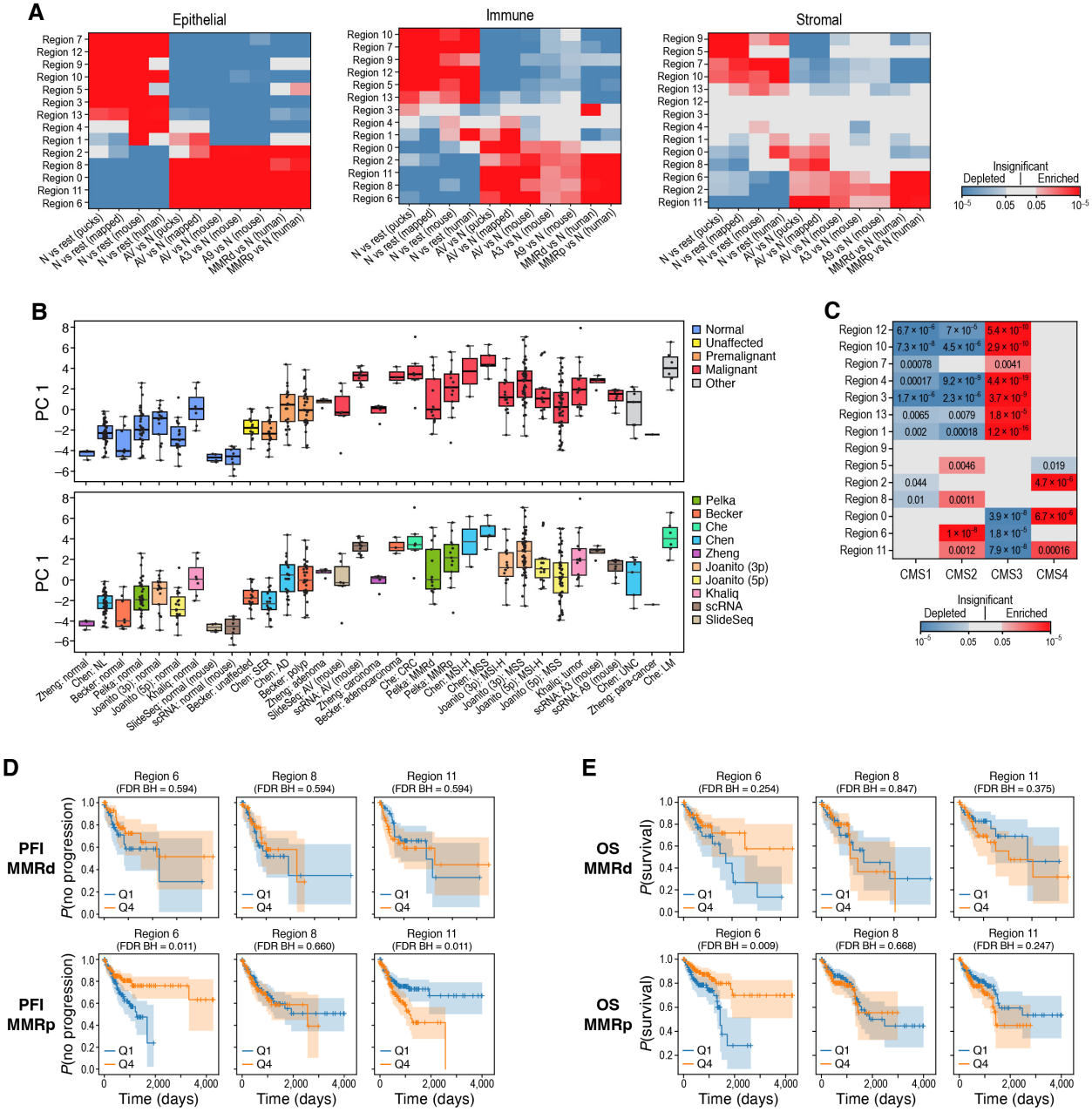

Composition
Cont7
Cont8
Cont9
Cont10
AV10a
AV10b
AV11a
AV11b
AV12
AV13
Region 13
Region 12
Region 11
Region 10
Region 9
Region 8
Region 7
Region 6
Region 5
Region 4
Region 3
Region 2
Region 1
Region 0
C
Normal
AV
Type composition
Distance from Region 2 (μm)
D
AV
Distance from Region 6 (μm)
Distance from Region 8 (μm)
Distance from Region 11 (μm)
E
Epithelial
B cell
T/NK cell
Monocyte
Macrophage
Granulocyte
Mast cell
Endothelial
Fibroblast
Sparc
Ctss
Mmp12
Cxcl9
Chd5
Mcam
Ctsd
Fraction of beads in group (%)
Mean expression in group
F
Region 6
Region 8
Region 11
Vim
Prox1
Sox11
G
Region 6 (d < 20 μm; P = 10)
Region 8 (d < 20 μm; P = 10)
Region 11 (d < 20 μm; P = 10)
z-score
H
Center gene = Cdh5
Center gene = Pecam1
Center gene = Pdgfrb
Cdh5
Pecam1
Pdgfrb
Epcam
Cd68
Jchain
Muc2
Cd3e
Co-occurrence: log2 (p(anno|center) / p(anno))
Distance (μm)
I
Reciever in Region 2 (FDR)
Sender in Region 2 (FDR)
Reciever in Region 6 (FDR)
Sender in Region 6 (FDR)
Reciever in Region 11 (FDR)
Sender in Region 11 (FDR)
enriched
depleted
PTN
PSAP
TAC
FGF
ANGPTL
GALECTIN
CX3C
NRG
GRN
GUCA
HH
MK
TNF
CD137
TRAIL
GCG
OSM
VEGI
VIP

1761 **Figure S8. Cellular neighborhoods.**

1762 **A.** Distinct characteristic region distributions in healthy and AV tissue. Proportion of UMIs
1763 assigned to each region (y axis) in each Slide-seq puck (x axis). **B.** β-catenin expression in AV
1764 lesions. Representative images (i) of β-catenin IHC on sections of AV lesions showing either
1765 nuclear (*) or cytoplasmic (**) staining pattern (x10 magnification), along with high magnification
1766 (x20, ii) of the region highlighted in (i), and H&E staining (iii) of a serial section showing a region
1767 proximal to the region in (i). (M-muscularis, L-lumen, dashed line: border between areas of nuclear
1768 and cytoplasmic staining of β-catenin). **C,D.** Spatial relations between the regions. Proportion of
1769 beads (y axis) of each region category (color code) at different distances (x axis) from region 2 (C,
1770 muscularis), 6 (D, left), 8 (D, middle) or 11 (D, right). **E,F.** Fraction of expressing cells (dot size)
1771 and mean expression per celltype (E, dot color) or per region (F, dot color) of marker genes
1772 (columns) across cell types in region 6 (E, rows) or regions 6, 8 and 11 (F, rows). **G.** Cell type
1773 neighborships in different malignant regions. Short-range (≤20μm) neighborship enrichment z-
1774 scores (color) *vs*. a background of spatially random annotation assignment for each pair of cell
1775 type annotations (rows, columns) in malignant-like regions 6, 8 and 11 within AV lesions. **H.**
1776 Spatial organization of endothelial cells and pericytes based on Cartana multiplex *in situ* RNA
1777 analysis. Co-occurrence (y axis) of Cdh5 (top) or Pecam1 (middle) -expressing (endothelial) cells
1778 or Pdgfrb (bottom) -expressing cells (pericytes) at different distances (x axis) from each central
1779 cell type (color code), based on expression of a specific gene (color bar). **I.** Cell-cell interaction
1780 pathways enriched in different regions. Enrichment (FDR) in regions 2 (muscularis, left), 6
1781 (malignant-like, inflammation, middle) or 11 (malignant-like, EMT, right) of corresponding
1782 "sender" (x axis) and "receiver" (y axis) (aggregated over ligand-receptor pairs) pathways (dots).
1783 Red/blue: pathways significantly enriched/depleted in region (light red/blue: only

1784 enriched/depleted for either sender or receiver). The top 5 enriched pathways (in each direction)

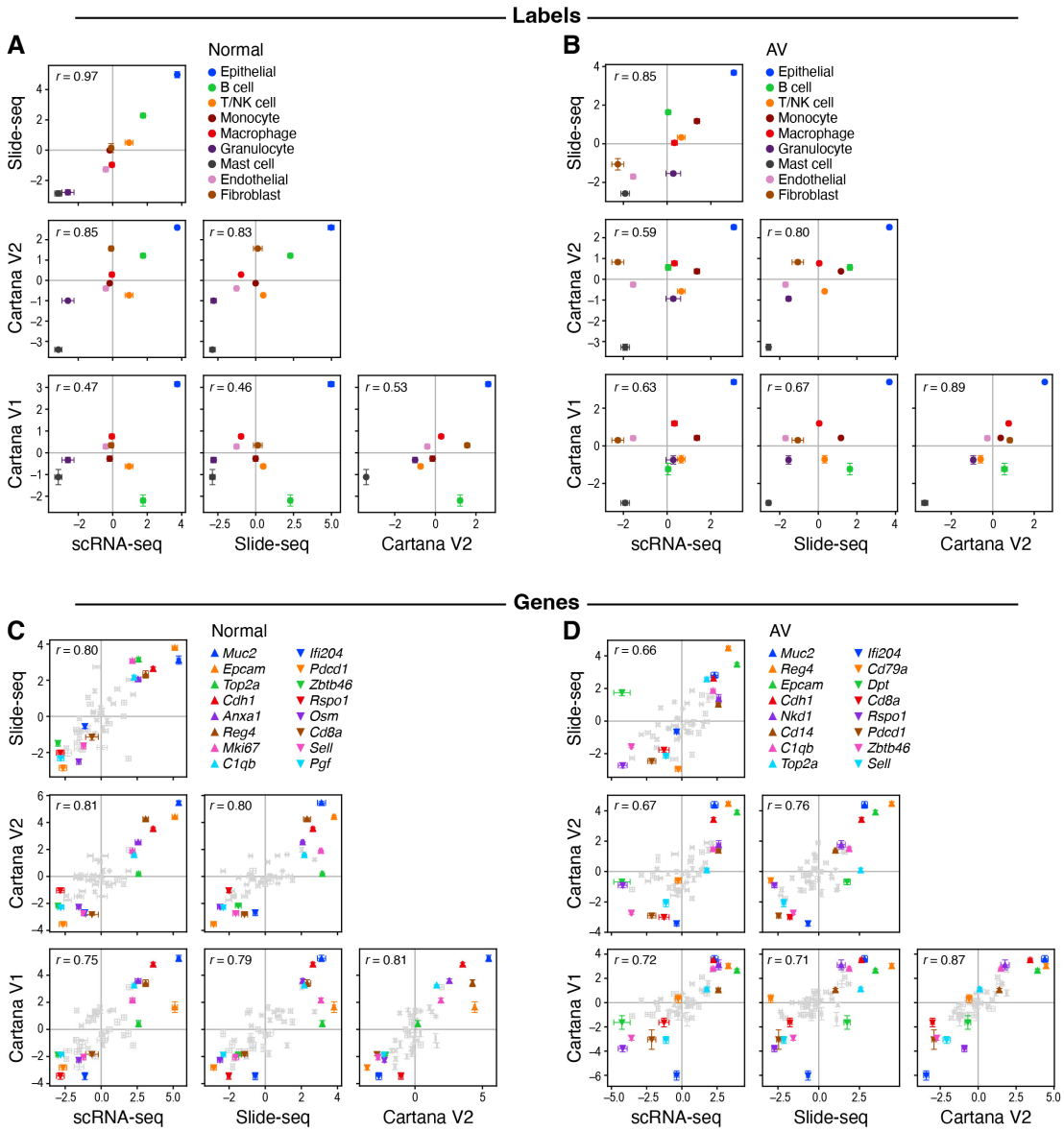
1785 are labeled.


A
Epithelial
Immune
Stromal
Insignificant
Depleted
Enriched
B
PC 1
Normal
Unaffected
Premalignant
Malignant
Other
Pelka
Becker
Che
Chen
Zheng
Joanito (3p)
Joanito (5p)
Khaliq
scRNA
SlideSeq
C
CMS1
CMS2
CMS3
CMS4
D
PFI
MMRd
MMRp
P(no progression)
Region 6
(FDR BH = 0.594)
Region 8
(FDR BH = 0.594)
Region 11
(FDR BH = 0.594)
Region 6
(FDR BH = 0.011)
Region 8
(FDR BH = 0.660)
Region 11
(FDR BH = 0.011)
Q1
Q4
Time (days)
E
OS
MMRd
MMRp
P(survival)
Region 6
(FDR BH = 0.254)
Region 8
(FDR BH = 0.847)
Region 11
(FDR BH = 0.375)
Region 6
(FDR BH = 0.009)
Region 8
(FDR BH = 0.668)
Region 11
(FDR BH = 0.247)
Q1
Q4
Time (days)

1793 **Figure S9. Transfer of mouse spatial region expression profiles to human patient data.**

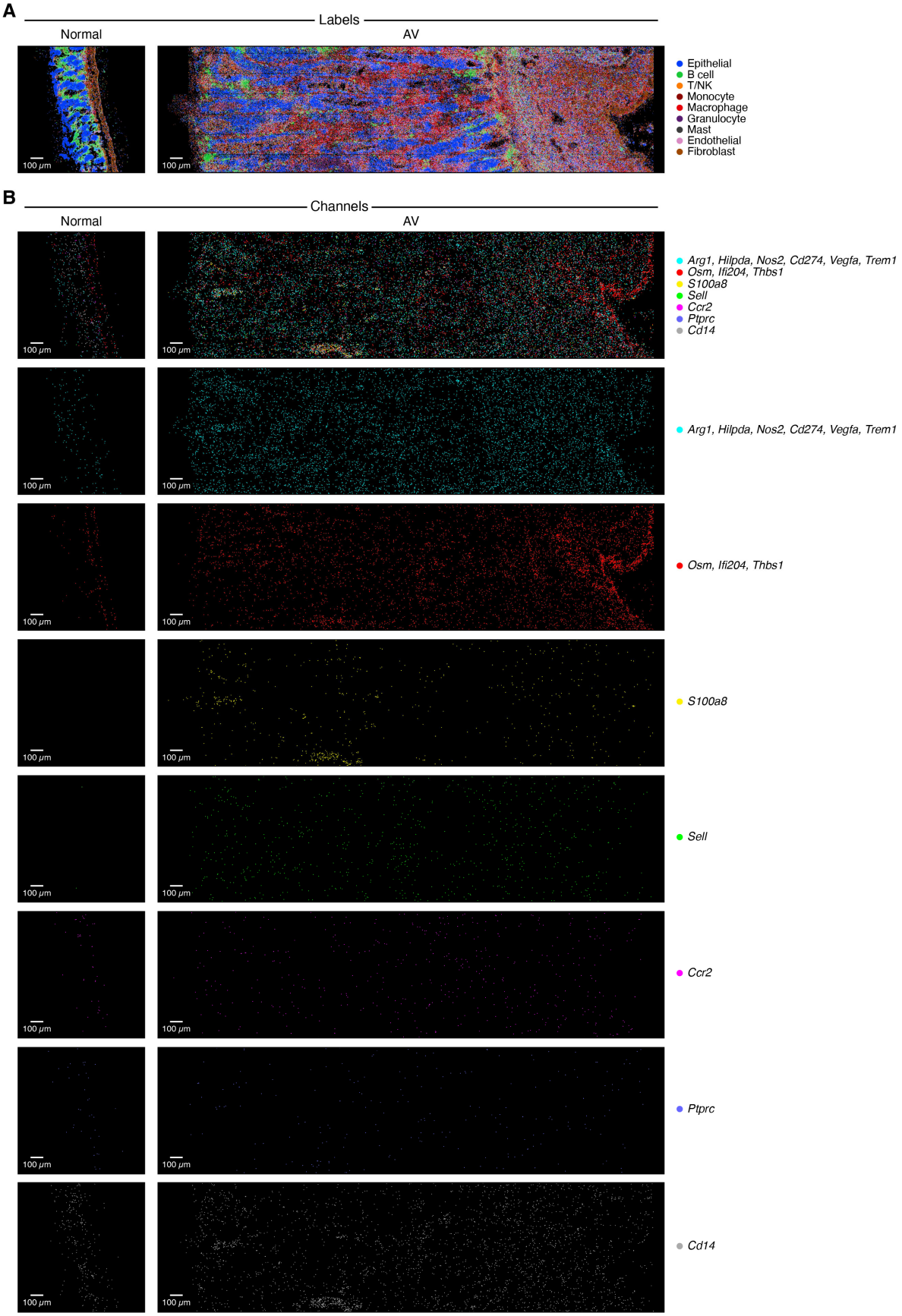
1794 **A.** Expression profiles characterizing mouse regions in human and mouse samples. Significance
1795 (FDR, color bar, two-sided Welch's t test on CLR-transformed compositions) of enrichment (red)
1796 or depletion (blue) of mouse region-associated epithelial (left), immune (middle) or stromal
1797 profiles (right) in pucks from normal colon and dysplastic lesions, in the same pucks but after
1798 mapping the region annotation to itself (consistency check), in mouse single cell data after
1799 mapping region annotation from mouse pucks, and in human single cell data after mapping region
1800 annotation from mouse pucks. **B**. PC1 values (y axis; box plots show mean, quartiles, and whiskers
1801 for the full data distribution except for outliers outside 1.5 times the interquartile range (IQR)) in
1802 a PCA of region scores of mouse and human samples (x axis) sorted by malignant status and
1803 colored by status (top, legend) or study (bottom, legend). **C**. Spatial regions profiles associated
1804 with different CMS classes. Significance (Benjamini-Hochberg FDR, color bar, two-sided
1805 Welch's t test) of enrichment (red) or depletion (blue) of each region profile (rows) in pseudo bulk
1806 profiles of human tumor samples classified in each CMS class (columns). **D,E**. Expression of
1807 malignant-like regions 6 and 11 in tumors is associated with PFI and OS in MMRd and MMRp
1808 patients. Kaplan-Meier PFI (f, MMRd n = 189^{95}, MMRp n = 447^{95}) or OS (g, MMRd n = 189^{95},
1809 MMRp n = 447^{95}) analysis of human bulk RNA-seq cohort stratified by malignant-like region
1810 profile scores.

1811


Labels
A
Normal
Epithelial
B cell
T/NK cell
Monocyte
Macrophage
Granulocyte
Mast cell
Endothelial
Fibroblast
r = 0.97
r = 0.85
r = 0.83
r = 0.47
r = 0.46
r = 0.53
Slide-seq
Cartana V2
Cartana V1
scRNA-seq
Slide-seq
Cartana V2
B
AV
r = 0.85
r = 0.59
r = 0.80
r = 0.63
r = 0.67
r = 0.89
Genes
C
Normal
Muc2
Epcam
Top2a
Cdh1
Anxa1
Reg4
Mki67
C1qb
Ifi204
Pdcd1
Zbtb46
Rspo1
Osm
Cd8a
Sell
Pgf
r = 0.80
r = 0.81
r = 0.80
r = 0.75
r = 0.79
r = 0.81
D
AV
Muc2
Reg4
Epcam
Cdh1
Nkd1
Cd14
C1qb
Top2a
Ifi204
Cd79a
Dpt
Cd8a
Rspo1
Pdcd1
Zbtb46
Sell
r = 0.66
r = 0.67
r = 0.76
r = 0.72
r = 0.71
r = 0.87

1820 **Supplementary item 1. Correspondence between scRNA-seq, Slide-seq, and Cartana data**

1821 **A,B**. Agreement in cell type proportions. Inferred cell type proportions (y and x axis; CLR-
1822 transformed cell type fractions) for each cell type (dot color) in normal (A) and AV (B) samples
1823 in scRNA-seq, Slide-Seq and Cartana data (axis labels). Pearson's r: top left corners. Error bars:
1824 bootstrapped standard error of the mean ((A) n=8, 16, 12, 16 and (B) n=11, 24, 12, 12 for scRNA-
1825 seq, Slide-seq, Cartana V1, Cartana V2). **C,D**. Agreement in gene expression levels. Measured
1826 expression (y and x axis; CLR-transformed gene fractions) for each gene (dot) measured by all
1827 three methods in normal (A) and AV (B) samples in scRNA-seq, Slide-Seq and Cartana data (axis
1828 labels). Pearson's r: top left corners. Error bars: bootstrapped standard error of the mean ((C) n=8,
1829 16, 12, 16 and (D) n=11, 24, 12, 12 for scRNA-seq, Slide-seq, Cartana V1, Cartana V2). Genes
1830 with the 8 maximal and minimal expression ranks across methods are labeled.

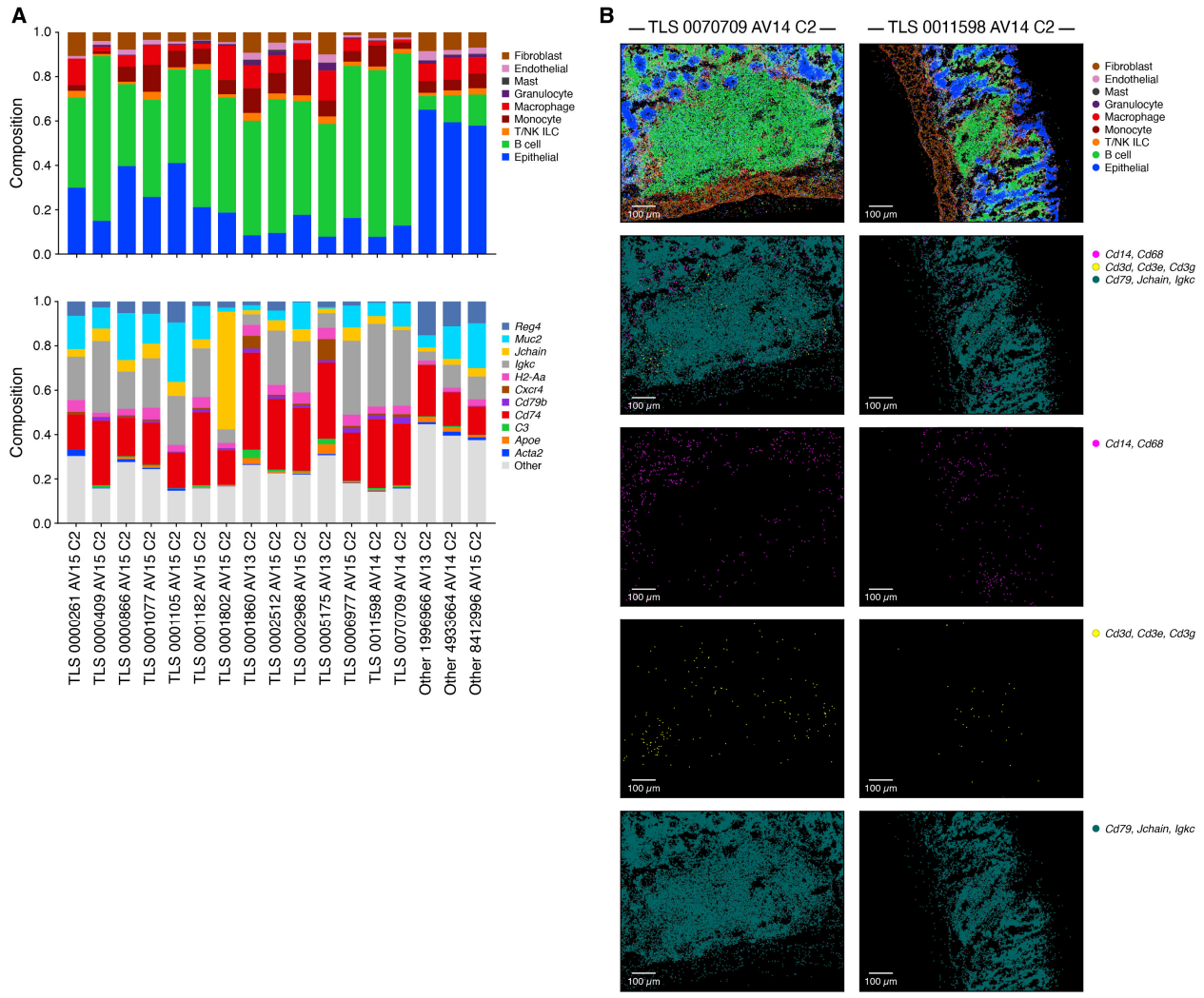

A
Labels
Normal
AV
Epithelial
B cell
T/NK
Monocyte
Macrophage
Granulocyte
Mast
Endothelial
Fibroblast
B
Channels
Normal
AV
Arg1, Hilpda, Nos2, Cd274, Vegfa, Trem1
Osm, Ifi204, Thbs1
S100a8
Sell
Ccr2
Ptprc
Cd14
100 μm

1831

1832 **Supplementary item 2. Dysplasia-associated cells in the tumor microenvironment.**

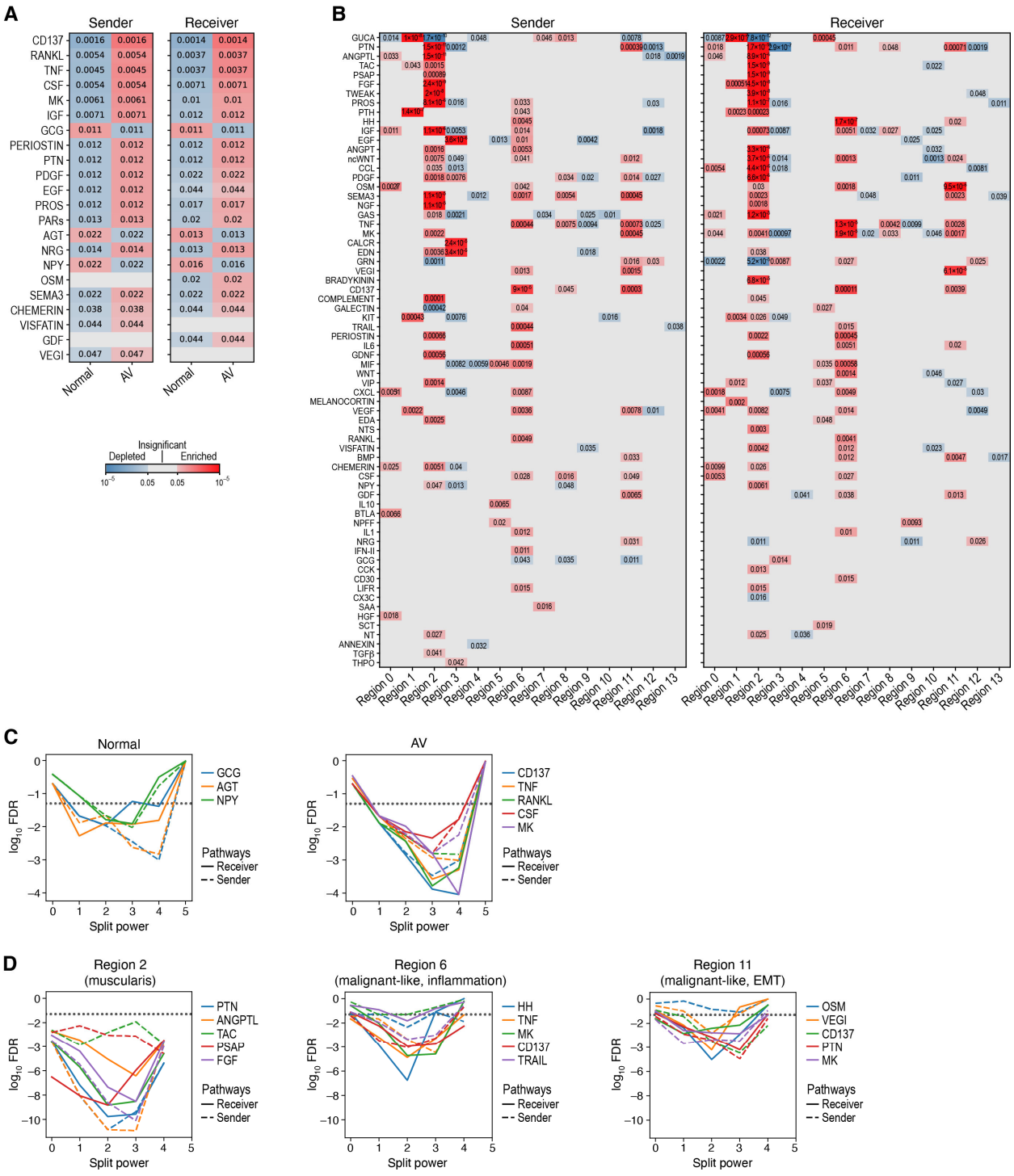
1833 **A,B.** Representative images of multiplex RNA analysis in normal colon (left) and AV lesions
1834 (right) colored by cell type assignment (A, the same is **Figure 1E**) and expression of marker genes
1835 (B) for Mono02 (Arg1, Hilpda, Nos2, Cd274, Vegfa, Trem1; cyan), Mono03 (Osm, Ifi204 and
1836 Thbs1; red), granulocytes (S100a8; yellow), as well as Sell (green), Ccr2 (magenta), Ptprc (blue)
1837 and Cd14 (gray).

1838


A
Composition
1.0
0.8
0.6
0.4
0.2
0.0
Fibroblast
Endothelial
Mast
Granulocyte
Macrophage
Monocyte
T/NK ILC
B cell
Epithelial
Reg4
Muc2
Jchain
Igkc
H2-Aa
Cxcr4
Cd79b
Cd74
C3
Apoe
Acta2
Other
TLS 0000261 AV15 C2
TLS 0000409 AV15 C2
TLS 0000866 AV15 C2
TLS 0001077 AV15 C2
TLS 0001105 AV15 C2
TLS 0001182 AV15 C2
TLS 0001802 AV15 C2
TLS 0001860 AV13 C2
TLS 0002512 AV15 C2
TLS 0002968 AV15 C2
TLS 0005175 AV13 C2
TLS 0006977 AV15 C2
TLS 0011598 AV14 C2
TLS 0070709 AV14 C2
Other 1996966 AV13 C2
Other 4933664 AV14 C2
Other 8412996 AV15 C2
B
TLS 0070709 AV14 C2
TLS 0011598 AV14 C2
100 μm
Cd14, Cd68
Cd3d, Cd3e, Cd3g
Cd79, Jchain, Igkc

1848 **Supplementary item 3. TLS-like structures in** AV **lesions by Cartana multiplex *in situ* RNA**
1849 **profiles.**

1850 **A.** Proportion of RNA molecules (y axis) attributed to each cell type (top) or gene (bottom) in
1851 TLS-like regions (x axis) in Cartana V2 multiplex *in situ* RNA experiment. Genes which constitute
1852 at least 3% of all measured transcripts in any of the identified follicular structures are noted
1853 (bottom). **B.** Representative images from TLS-like regions showing the distribution of cell type
1854 (top row) and of cell type markers (bottom four rows; Cd14, Cd68 (monocytes, magenta), Cd3d,
1855 Cd3e, Cd3g (T cells, yellow), and Cd79, Jchain, Igkc (B cells, dark cyan)).


A
Sender
Receiver
Normal
AV
Insignificant
Depleted
Enriched
B
Sender
Receiver
C
Normal
AV
log10 FDR
Split power
Pathways
Receiver
Sender
GCG
AGT
NPY
CD137
TNF
RANKL
CSF
MK
D
Region 2
(muscularis)
Region 6
(malignant-like, inflammation)
Region 11
(malignant-like, EMT)
PTN
ANGPTL
TAC
PSAP
FGF
HH
TRAIL
OSM
VEGI

1860 **Supplementary item 4. Ligand-receptor analysis.**

1861 **A,B.** Cell-cell interaction pathways enriched in normal and AV tissue and regions. Significance
1862 (FDR, color bar, two-sided Mann-Whitney U test) of enrichment (red) or depletion (blue) of
1863 expression of "sender" (left) and "receiver" (right) cell-cell communication pathways (rows)
1864 (aggregted from ligand-receptor pairs) in normal *vs*. AV samples (A, columns), or regions (B,
1865 columns). **C,D.** Split power analysis for cell-cell communication pathway analysis. Significance
1866 of enrichment ($\log_{10}$(FDR), y-axis) for receiver (solid lines) and sender (dashed line) for each of
1867 the top 5 enriched communication pathways (color) in each iteration for the spatial split of the
1868 Slide-Seq pucks into spatially disconnected and separated patches (x axis), for either normal vs.
1869 AV tissue enrichment (C) or for specific regions (D).