CDH1 loss promotes diffuse-type gastric cancer tumorigenesis via epigenetic reprogramming and immune evasion

Gengyi Zou,[1,†] Yuanjian Huang,[1,2,†] Shengzhe Zhang,[1] Kyung-Pil Ko,[1] Bong Jun Kim,[1] Jie Zhang,[1] Melissa P. Pizzi,[3] Yibo Fan,[3] Sohee Jun,[1] Na Niu,[6] Huamin Wang,[7] Shumei Song,[3] Jaffer A. Ajani,[3] Jae-Il Park[1,4,5]

[1]Department of Experimental Radiation Oncology, Division of Radiation Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA

[2]Department of General Surgery, The First Affiliated Hospital of Nanjing Medical University, Nanjing, Jiangsu 210029, China

[3]Department of GI Medical Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA

[4]Graduate School of Biomedical Sciences, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA

[5]Program in Genetics and Epigenetics, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA

[6]Department of Pathology, Yale School of Medicine, New Haven, CT 06510, USA

[7]Department of Pathology, Division of Pathology/Lab Medicine, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA

[†]These authors contributed equally.

[‡]Correspondence: Jae-Il Park (jaeil@mdanderson.org)

Tel: 713-792-3659; Fax: 713-794-5369

Keywords: CDH1, E-Cadherin, gastric cancer, gastric organoids, diffuse-type gastric adenocarcinoma, immune landscape remodeling, immune evasion, single-cell transcriptomics, cancer subtyping, EZH2

Abstract

Diffuse-type gastric adenocarcinoma (DGAC) is lethal cancer often diagnosed late and resistant to therapeutics. Although hereditary DGAC is mainly characterized by mutations in the *CDH1* gene encoding E-cadherin, the impact of E-cadherin inactivation on sporadic DGAC tumorigenesis remains elusive. We found that CDH1 inactivation occurs only subset of DGAC patient tumors. Unsupervised clustering of single-cell transcriptomes of DGAC patient tumors identified two subtypes of DGACs: DGAC1 and DGAC2. The DGAC1 is mainly characterized by CDH1 loss and exhibits distinct molecular signatures and aberrantly activated DGAC-related pathways. Unlike DGAC2 lacking immune cell infiltration in tumors, DGAC1 tumor is enriched with exhausted T cells. To demonstrate the role of CDH1 loss in DGAC tumorigenesis, we established a genetically engineered murine gastric organoid (GOs; *Cdh1* knock-out [KO], *Kras*^{G12D}, *Trp53* KO [EKP]) model recapitulating human DGAC. In conjunction with *Kras*^{G12D}, *Trp53* KO (KP), *Cdh1* KO is sufficient to induce aberrant cell plasticity, hyperplasia, accelerated tumorigenesis, and immune evasion. Additionally, EZH2 was identified as a key regulon promoting CDH1 loss-associated DGAC tumorigenesis. These findings underscore the significance of comprehending the molecular heterogeneity of DGAC and its potential implication for personalized medicine to DGAC patients with CDH1 inactivation.

Introduction

Gastric adenocarcinoma (GAC) is the 4th most common cause of cancer deaths worldwide [1]. GAC is mainly divided into intestinal-type gastric adenocarcinoma (IGAC, 50%), diffuse-type gastric adenocarcinoma (DGAC, 30%), and mixed [2]. DGAC is histologically characterized by poor differentiation, loss of cell adhesion proteins, fibrosis, and infiltration. Unlike IGAC, DGAC is relatively more often observed in younger, female, and Hispanic population than in older, male, and non-Hispanic ones [3-6]. While the incidence of IGAC has declined due to H. Pylori (HP) therapy and lifestyle improvements over the past few decades, the number of DGAC cases has remained constant or has risen[7,8].

DGAC tends to metastasize to the peritoneal cavity, which makes it difficult to diagnose early by imaging. In addition, isolated tumor cells or small clusters of tumor cells infiltrate in unpredictable patterns. Thus, DGAC is often detected at a late stage, leading to a poor prognosis. For such patients, curative resection is not possible. Systemic therapy is the main option for potentially prolonging survival and improving symptoms [9,10]. Despite the distinct features of DGAC in both a molecular basis and therapy resistance, the first-line treatment options are not specific for DGAC[11-13]. Systemic therapy with targeted therapy has shown limited benefits[14,15]. In parallel, immune checkpoint inhibitors (ICIs) have been used recently. The advent of first-generation ICIs that target Cytotoxic T-Lymphocyte Antigen 4 (CTLA4) and

69 Programmed death-ligand (PD-L1) has brought a paradigm shift in the treatment of
70 various advanced cancers [16]. Nivolumab (PD-1 inhibitor) can be either combined with
71 chemotherapy as first-line treatment or used as monotherapy as later-line treatment in
72 Asia [17,18]. Pembrolizumab (PD-1 inhibitor) showed a promising outcome treating GAC
73 with high microsatellite instability or high tumor mutational burden[19]. However, DGAC
74 imposes major difficulty in the clinic and available therapies perform poorly. Generally,
75 DGAC has immunosuppressed stroma and is genomically stable [20,21]. Tumor
76 microenvironment (TME) of DGAC often expresses the second generation of
77 checkpoints such as T-cell immunoglobulin mucin receptor 3 (TIM3) and its ligand
78 galectin-9, which induce immune landscape remodeling for immune evasion [22,23]. Given
79 the limited therapeutic options for DGAC, it is imperative to understand the biology of
80 DGAC, which may establish a groundwork for developing new targeted therapies for
81 DGAC. Furthermore, for maximizing therapeutic efficacy, it is crucial to identify patients
82 who can most benefit from specific treatment options. Nevertheless, to date, DGAC
83 patient stratification by molecular signatures has not been achieved.

84 Hereditary DGAC, as a minor proportion of DGAC (1–3%), is mainly
85 characterized by germline mutations in the *CDH1* gene that encodes E-cadherin [24].
86 However, other than Hereditary DGAC, the role of CDH1 loss in DGAC tumorigenesis is
87 unclear. Cell-to-cell adhesion is a crucial phenomenon for maintaining tissue
88 morphogenesis and homeostasis, as well as for regulating cell differentiation, survival,
89 and migration. E-cadherin mediates cell-to-cell adhesion, which is essential for
90 determining the proliferation specificity and differentiation of epithelial cells and

91 preventing invasion [25]. To understand the impact of *CDH1* loss on DGAC
92 tumorigenesis, we analyzed single-cell transcriptomes of 20 DGAC patient tumor
93 samples and identified two subtypes of DGACs exhibiting specific molecular
94 signatures including E-cadherin loss and immune landscape remodeling. To further
95 verify our in-silico analysis, we generated and characterized a genetically engineered
96 gastric organoid model that recapitulates E-cadherin inactivation-associated DGAC
97 tumorigenesis. This study stratifies DGAC patients by single-cell transcriptomics, and
98 elucidates the unexpected role of E-cadherin loss in cell plasticity, transcriptional
99 reprogramming, and immune evasion, providing novel insights into E-cadherin loss-
100 associated DGAC tumorigenesis.

101

Results

CDH1 inactivation in DGAC

To explore the role of CDH1 in DGAC, we examined the genetic alterations, mRNA expression, and protein levels of CDH1 in DGAC. 25% of tumor cells from the DGAC patients showed *CDH1* gene alterations, including mutations and deep deletions (**Fig. 1A**). We also assessed the CDH1 protein expression in the tissue microarray of 114 DGAC patients' tumor samples (patient information was listed in **Table S4**). Immunohistochemistry (IHC) showed that 37.72% of DGAC patients were CDH1 negative, 37.72% exhibited low CDH1 expression, and 24.56% displayed high CDH1 expression (**Fig. 1B**, which was also quantified with histochemical scoring assessment (H-score) of each slide (**Fig. 1C**). Next, we determined the transcriptional signature of DGAC at the single-cell transcriptomics level by analyzing single-cell RNA-seq (scRNA-seq) datasets of 20 stage IV DGAC patients' tumor samples (**Fig. 1D, Table S5**) [26]. After data integration and normalization, a total of 27 cell clusters was generated according to distinctive gene expression patterns (**Fig. 1E, fig. S1A, B, Table S6**). We re-clustered the datasets as the mega clusters according to Leiden-based UMAP (**Fig. 1F**). To conduct the precise subtyping of DGAC, we reanalyzed the scRNA-seq datasets with only epithelial cells (**Fig. 1G, fig. S1C, Table S7**). An unsupervised pair-wise correlation analysis showed that the combined datasets of 20 DGAC patients were divided into two major subtypes (DGAC1 and DGAC2) (**Fig. 1H**). The

transcriptional signature of DGAC1 epithelial cell clusters was highly distinct from that of DGAC2 (**Fig. 1I, fig. S1D, Table S8**). In line with the heterogeneity of CDH1's genomic alterations and expression in DGAC patients (**Fig. 1A, B**), the DGAC1 subtype exhibited a significantly lower expression of *CDH1* compared to DGAC2 (**Fig. 1J, K**), indicating that the unsupervised pair-wise subtyping can also stratify DGAC patients by *CDH1* expression. We also identified the molecular signatures of DGAC1 and DGAC2 (**Fig. 1L**). The DGAC1 tumors were enriched with the expression of *TXNIP* (thioredoxin interacting protein), *EVL* (Ena/Vasp-Like), *TSC22D3* (TSC22 Domain Family Member 3; also known as glucocorticoid-induced leucine zipper, GILZ) genes (**Fig. 1L**). A high level of TXNIP expression is associated with significantly shorter survival of patients with non-small cell lung cancer and invasive growth of hepatocellular carcinoma [27,28]. It has also been reported that decreased TXNIP RNA expression is associated with poor prognosis of patients with clear cell renal cell carcinoma [29]. EVL belongs to the Ena/VASP (Enabled/vasodilator-stimulated phosphoprotein) family of proteins, which have a range of roles in regulating the actin cytoskeleton [30]. Studies has shown that EVL is upregulated in breast cancer [31]. Meanwhile, the upregulation of TSC22D3 can subvert therapy-induced anticancer immunosurveillance [32]. In addition, we also identified the molecular signatures of DGAC2 (*SPINK1*, *IFI27*, and *TSPAN8*) (**Fig. 1L**). These results identify two distinct subtypes of DGACs by distinct molecular signatures and CDH1 expression.

Molecular characterization of DGAC subtypes

Next, we characterized the molecular subtypes of DGAC. Given that CDH1 loss confers the epithelial-mesenchymal transition (EMT) process, we checked the EMT scores based on the established gene set (**Table S9**). DGAC1 showed a higher EMT score compared to DGAC2 (**Fig. 2A, fig. S2A**). Extensive genomic analyses of GAC have found that DGACs display distinct activation of signaling pathways different from IGACs [33]. scRNA-seq-based signaling scoring showed that FGFR2 and PI3K/AKT/mTOR pathways were activated in DGAC1 (**Fig. 2B, C, fig. S2B, C**), while RHOA, MAPK, HIPPO, WNT, and TGF-β pathways were activated in DGAC2 (**Fig. 2D-H, fig. S2D-H**). In addition, we analyzed the copy number variation (CNV) of DGACs by using normal stomach samples as a reference. We combined 29 scRNA-seq online datasets of normal stomach samples (Normal) with the previous 20 DGAC patients [34] (**Fig. 2I**). Except for the endothelial cell markers, the same marker panel was utilized as the previous DGAC subcategory process to annotate the cells into epithelial cells, myeloid cells, B cells, plasma cells, T cells, effector T cells, naïve T cells, exhausted T cells, fibroblasts, and endothelial cells (**fig. S1A, S3A**). Leiden-based UMAP exhibited the same cell types as the DGAC stratification analysis (**Fig. 2J, K, fig. S2B, Table S10**), except that the endothelial cell cluster appeared due to the normal tissue (**fig. S3A**). According to the previously identified DGAC subgroups, we separated the UMAP as Normal, DGAC1, and DGAC2 (**Fig. 2L, fig. S3C**). Although the epithelial cells were defined as EPCAM ^{high} clusters among all groups, epithelial cells from the Normal group were clearly isolated from the major epithelial cell population of the merged datasets (**Fig. 2K, L**). CNV patterns were somewhat different between DGAC1 and DGAC2 (**Fig.**

167 **2M**). The higher CNV scores were observed in DGACs compared to the Normal (**Fig.**
168 **2N, O**). These results indicate the heterogeneity of DGAC with differentially activated
169 signaling pathways.

170

171 ## Immune landscape remodeling with T cell exhaustion in DGAC1

172 Having determined the molecular signatures of DGAC tumor cells, we next analyzed
173 TME. Intriguingly, scRNA-seq-based immune cell profiling showed that compared to
174 DGAC2 where immune cells barely existed, DGAC1 was highly enriched with immune
175 cells, including T cells, B cells, and myeloid cells (**Fig. 3A-C, fig. S4**). Additionally, we
176 examined cellular networks among all cell clusters (DGAC1 vs. DGAC2) using a
177 CellChat package that infers cell-to-cell functional interaction based on ligand-receptor
178 expression [35]. Compared to DGAC2, DGAC1 showed relatively more inferred
179 interactions among different cell types (**Fig. 3D**). According to the differential number of
180 interactions, the interactions between fibroblast and epithelial and endothelial cells
181 were decreased, while widespread increased interactions were found in DGAC1
182 compared to DGAC2 (**Fig. 3E**). Notably, exhausted T cells, as a receiver, showed the
183 most increased interactions compared with other cell types in DGAC1 (**Fig. 3F**). fGSEA
184 (fast GeneSet Enrichment Analysis) identified the pathways that are enriched in DGAC1
185 with six gene sets, including GOBP (Gene sets derived from the Gene Ontology
186 Biological Process), and five canonical pathways gene sets (REACTOME, WP,
187 BIOCARTA, PID, and KEGG) (**fig. S5, S6**). Except for REACTOME (**fig. S5B**), T cell-

188 related immune response pathways were enriched in DGAC1 based on the other five
189 gene sets (**fig. S5A, C, S6A-C**). Consistent with the Cell Chat prediction and fGSEA
190 results, DGAC1 showed the significant upregulation of T cell exhaustion markers
191 (LAG3, TIGIT, CTLA4, and HAVCR2) and the increased T cell exhaustion score,
192 compared to DGAC2 (**Fig. 3G, I-K**). Similarly, immune checkpoints-related genes
193 (CTLA4, PDCD1, PDCD1LG2, and CD274) and their score were markedly upregulated
194 in DGAC1 over DGAC2 (**Fig. 3H, I, L, M**). In addition to T cell analysis, we also
195 examined myeloid-derived suppressor cells (MDSC) and macrophage polarization. We
196 observed MDSC score was also increased in DGAC1 compared to DGAC2 while no
197 obvious changes of macrophage polarization between DGAC1 and DGAC2 (**fig. S7**).
198 These results suggest that compared to DGAC2, the DGAC1 subtype exhibits distinct
199 immune remodeling featured by T cell exhaustion and increased expression of the
200 genes associated with immune checkpoints.

201

202 *Cdh1* KO induces hyperplasia in the murine GOs

203 To validate the in silico results, we utilized murine GOs that enable multiple genetic
204 engineering with immediate phenotype analyses. *Cdh1* deficiency results in early-stage
205 DGAC phenotype in a mouse model [36,37]. Nevertheless, other genes need to be
206 included to recapitulate DGAC tumorigenesis. The genes encoding the receptor
207 tyrosine kinase (RTK)-RAS signaling pathway and the *TP53* gene were profoundly
208 altered in DGAC [22,38]. *KRAS* and *TP53* were genetically altered in 13.19% and 36.11%

209 of DGAC cases, respectively, as per cBioportal analysis (**Fig. 4A**). Therefore, we
210 genetically manipulated three genes (*Cdh1*, *Trp53*, and *Kras*) in GOs. Briefly, from the
211 *Cdh1* wild type (WT) and *Kras* $^{LSL\text{-}G12D/+}$; *Trp53* $^{fl/fl}$ mice, gastric epithelial cells were
212 isolated to culture them into GOs (**Fig. 4B**). Subsequently, using the Cre-LoxP
213 recombination and CRISPR-based genetic manipulation, we established two lines of
214 GOs carrying *Kras* $^{G12D/+}$ and *Trp53* deletion in combination with *Cdh1* KO (KP:
215 *Kras* $^{G12D/+}$; *Trp53* KO [KP], *Cdh1*/*E-Cadherin* KO; *Kras* $^{G12D/+}$; *Trp53* KO [EKP]) (**Fig. 4B**).
216 Genetic modifications were validated by PCR-based genotyping and genomic DNA
217 sequencing and immunofluorescence (IF) staining (**fig. S8, Fig. 4G**). Meanwhile, we
218 monitored their sizes and numbers by macroscopic analyses during passages to
219 maintain the stable culture process during passages (**Fig. 4C, D**). Unlike WT GOs
220 growing as a single layer of epithelial cells, KP and EKP GOs displayed multilayered
221 epithelium (**Fig. 4E**). Notably, compared to WT and KP, EKP GOs exhibited abnormal
222 morphology such as vacuolization and cell adhesion loss along with cell hyperplasia
223 (**Fig. 4E**). Additionally, EKP GOs were hyperproliferative compared to WT and KP GOs,
224 assessed by immunostaining of MKI67, a cell proliferation marker (**Fig. 4F, H**). These
225 results suggest that in conjunction with *Trp53* KO and *Kras* ^{G12D}, *Cdh1* loss is sufficient
226 to induce hyperplasia.

227

228 **Cdh1 loss induces aberrant gastric epithelial cell plasticity**

229 We next interrogated the mechanism of Cdh1 loss-associated DGAC tumorigenesis by
230 multiplex scRNA-seq of WT, KP, and EKP GOs (**fig. S9A**). Each group was tagged with
231 two CMO (Cell Multiplexing Oligo) tags, then pooled together with the same number of
232 cells after being counted. All datasets were integrated with the Harmony algorithm [39] to
233 minimize the batch effect (**fig. S9B**). WT, KP, and EKP GOs were merged well in a
234 batch-based UMAP (**Fig. 5A**). To identify the gene signature of each cell cluster, we
235 generated a heatmap to calculate the top 5,000 highly variable genes (**fig. S9C**). Each
236 UMAP and heatmap represented the different cell distribution among three types of
237 GOs (**Fig. 5B, C, fig. S9D-F**) with distinct marker gene expression shown in the dot
238 plot (**Fig. 5D, Table S11**). Notably, Aquaporin 5 (Aqp5), a gastric tissue stem cell
239 marker [40], was decreased in EKP compared to WT and KP (**Fig. 5C**). Next, we
240 determined the pathological relevance of GO models to human GAC by assessing the
241 expression of genes related to Mucins, cell stemness, and clinical GAC markers. The
242 dot plots showed that compared to KP GOs, mucinous markers (especially, *Muc1*)
243 were highly upregulated in the EKP GOs (**Fig. 5E, I**). Consistent with cell proportion
244 results, the EKP GOs showed a relatively higher expression of *Mki67* compared to WT
245 and KP (**Fig. 5F, K, L**). The pathological diagnostic markers of human GAC include
246 KRT7, KRT20, and CDX2 (GAC markers); MUC1, MUC2, MUC5AC (diagnostic
247 differentiation markers); SOX2 and SOX4 (undifferentiation or stemness markers).
248 Among all panels, the expressions of *Krt7*, *Muc1*, and *Sox4* were markedly increased in
249 EKP GOs compared to other GOs (**Fig. 5G, J, fig. S10**).

To determine the pathological relevance of EKP GOs with human DGAC, we utilized a single-cell inferred site-specific omics resource (Scissor) analysis [41] and assessed the transcriptomic similarity between of EKP GOs and the bulk RNA-seq data of patients diagnosed with DGAC from the TCGA database. While as a reference, the transcriptional signature of WT GOs was matched with that of normal stomach tissue, EKP GOs displayed similar transcriptional features to that of DGAC (**Fig. 5H**), indicating that EKP GOs are similar to the subtype of human DGAC at the level of gene expression.

Having observed the significant impact of Cdh1 loss on hyperplasia (Mki67+ cell cluster) and gastric tissue stem cell marker expression (Aqp5+ cell cluster), we investigated the cellular mechanism of cell transformation provoked by *Cdh1* loss. We analyzed WT, KP, and EKP GO scRNA-seq datasets for the cell lineage trajectory inference by using the CytoTRACE algorithm [42]. While Aqp5^{high} cell cluster served as a cellular origin in WT and KP GOs, Miki67^{high} cells became the primary cellular origin of EKP GOs (**Fig. 5M-O**), which was consistent with the cell proportion results (**Fig. 5C**). In addition to *Mki67*, *Hmgb2l*, and *Pclaf*, additional markers for proliferating cells were significantly increased in the proliferating cell clusters of EKP GOs, compared to those of KP GOs (**Fig. 5P**). These results suggest that CDH1 inactivation is sufficient to induce aberrant cell lineage commitment with the generation of the distinct hyperplastic cellular origin.

271 ## *Cdh1* KO induces immune evasion of tumor cells

272 Having determined distinct immune remodeling with T cell exhaustion in the DGAC1
273 subtype where CDH1 is downregulated (**Fig. 3**), we asked whether genetic ablation of
274 *CDH1* contributes to immune evasion of DGAC. To test this, we established KP and
275 EKP GO-derived cell lines in 2D culture with minimum growth factors (culture medium:
276 DMEM Complete Medium with 10% fetal bovine serum and 1% penicillin-
277 streptomycin) for allograft transplantation (**Fig. 6A**). Unlike WT GOs that failed to grow
278 in 2D culture, both KP and EKP cells grew in 2D culture and were maintained well at
279 multiple passages. Then, KP and EKP cell lines derived from C57BL/6 strain were used
280 for transplantation into C57BL/6 mice. The morphological characteristics of KP and
281 EKP cells exhibited notable differences. KP cells exhibited a compact and tightly
282 packed phenotype, forming densely clustered colonies, while EKP cells displayed a
283 more loosely-arranged and dispersed morphology, lacking the cohesive structure of
284 KP cells (**Fig. 6B**). Of note, there was no significant difference in cell proliferation
285 between KP and EKP cells (**Fig. 6C**). However, transplantation results showed that
286 tumor incidence and volume of EKP tumors was markedly higher than KP tumors
287 (tumor incidence rates: EKP [91.7%] vs. KP [16.7%]) (**Fig. 6D-F**). Histologically, EKP
288 tumors exhibited poorly differentiated tumor cells, the feature of DGAC (**Fig. 6G**) with
289 increased cell proliferation (**Fig. 6H, M**) and CDH1 loss (**Fig. 6I**). Compared to KP
290 tumors, EKP tumors showed relatively increased numbers of immune cells expressing
291 PDCD-1 and TIM3 (also called HAVCR2), while cells expressing CD3, a marker for T
292 cells, remained similar (**Fig. 6J-L, N-P**). These results suggest that CDH1 is a

293 gatekeeper restricting the immune evasion of DGAC, confirming immune landscape
294 remodeling associated with the DGAC1 subtype where CDH1 is inactivated.

295

296 *Cdh1* depletion-activated EZH2 regulon promotes gastric tumorigenesis

297 Since CDH1 loss induced cell lineage plasticity and transcriptional reprogramming, we
298 sought to identify key transcriptional regulatory modules (regulons) activated by *Cdh1*
299 depletion. We integrated the scRNA-seq datasets of WT, KP, and EKP into batch-
300 based and regulon pattern-based UMAPs (**Fig. 7A**). In the regulon activity-based
301 UMAP, six major transcriptional clusters (0~5) were identified (**Fig. 7A**). With the
302 separated UMAP, we observed that WT and KP shared somewhat similar
303 transcriptional landscape. However, EKP exhibited distinct features with an increased
304 cluster 5 (**Fig. 7B**). To pinpoint essential regulons, we created an unbiased workflow
305 (**Fig. 7C**). Based on the Z score of each regulon, we identified 32 regulons specific to
306 EKP transcriptional profile, compared to those of WT and KO (**Fig. 7D**). Additionally,
307 regulon specificity score (RSS) analysis showed the top 20 regulons specific to EKP
308 (**Fig. 7E**). RSS-based top 20 regulons belonged to Z score-based regulons (**Fig. 7F,**
309 **Table S12**). Both RSS and Z-score were used to quantify the activity of a gene or set
310 of genes. Z-score was used to quantify the level of gene expression in a particular
311 sample, while RSS was used to quantify the specificity of a gene set to a particular
312 regulatory network or module [43]. According to TCGA-based upregulation in DGAC
313 patients compared to normal stomach tissues, 13 regulons (Brca1, E2f1, E2f3, E2f7,

314 E2f8, Ezh2, Gabpa, Gtf2b, Gtf2f1, Hmga2, Pole4, Sox4, and Tfdp1) were selected (**fig.**
315 **S11A**). Next, we examined the regulons' expression in organoids datasets. Compared
316 to WT and KP, the expression of Ezh2, Gtf2b, Pole4, and Sox4 was obviously
317 increased in EKP GOs with over 40% fractions of clusters (**Fig. 7G**). According to the
318 regulon activity-based UMAP, Ezh2 displayed the highest score in EKP compared to
319 WT and KP GOs (**Fig. 7H, fig. S11B**). To assess the pathological relevance of EZH2 to
320 DGAC, we analyzed the expression of downstream target genes of EZH2 in the DGAC
321 datasets (**Table S9**). Compared to DGAC2 (CDH1 high), the EZH2 target gene score
322 was indeed relatively higher in DGAC1 (CDH1 loss) (**Fig. 7I, J**). EZH2 is a histone
323 methyltransferase catalyzing the methylation of histone H3 lysine 27 (H3K27) to
324 generate H3K27me3, which is associated with gene repression [44]. Consistent with
325 EZH2 regulon activation by *Cdh1* KO, H3K27Me3 was also increased in EKP tumors
326 compared to KP, while no significant difference in H3K27Ac expression (**Fig. 7K**). Next,
327 we treated EKP cells with GSK343, a specific inhibitor of EZH2 methyltransferase [45].
328 EKP cells were more sensitive to GSK343 compared with KP for in vitro cell growth
329 (**Fig. 7L**). Additionally, allograft transplantation experiments showed the growth
330 inhibitory effect of GSK343 on EKP tumorigenesis (**Fig. 7M-O**). These results identify
331 EZH2 as a key regulon contributing to tumorigenesis of CDH1 inactivation-associated
332 DGAC.

333

Discussion

The impact of CDH1 loss on sporadic DGAC tumorigenesis remains unknown. Single-cell transcriptomics-based unsupervised clustering identified two subtypes of DGAC: DGAC1 (CDH1-negative or downregulated) and DGAC2 (CDH1-positive). Unlike DGAC2 lacking tumor-infiltrated immune cells, the DGCA1 subtype is enriched with exhausted T cells. Single-cell transcriptomics and transplantation assays showed that *Cdh1* KO induces aberrant cell plasticity, hyperplasia, accelerated tumorigenesis, and immune evasion. Moreover, EZH2 regulon specifically activated by CDH1 loss promotes DGAC tumorigenesis.

Patient stratification is crucial for improving therapeutic efficacy. Despite several studies classifying GAC patients [21,46-49], such subtyping did not consider single-cell level cellular convolution, which might be insufficient to represent the full spectrum of DGAC features. Our stratification approach was based on the high dimensional transcriptional signatures at the single-cell level, immune cell profiling, and cellular network, which may complement limitations from the bulk analyses and likely better stratify DGAC patients. Indeed, our unsupervised subtyping by tumor cell transcriptome well matched with distinct immune cell properties (**Fig. 3A-C**). Furthermore, the application of CellChat and fGSEA analysis led to the identification of T cell-related immune profiling as the dominant feature in DGAC1 (**Fig. 3D-F, fig. S5, S6**). Interestingly, T cell exhaustion and immune checkpoint-related genes were notably enriched in DGAC1 compared to DGAC2 (**Fig. 3G-M**), confirmed by the transplantation experiments (**Fig. 6**). These results strongly suggest that DGAC1 patients might benefit from T cell-based ICIs. Conversely, DGAC2 patients might be ICI non-responders since T cells barely exist in the tumors (**Fig. 3**).

Understanding the biology of cancer immune evasion is also imperative for improving cancer treatment. To date, how DGAC tumor cells evade immune surveillance remains elusive. Transplantation assays showed that CDH1 loss is

364 sufficient for immune evasion of DGAC (**Fig. 6**). In line with this, EKP allografts
365 displayed increased expression of PDCD1 and TIM3 (**Fig. 6K-L**), also identified as
366 molecular signatures of DGAC1 (**Fig. 3G-M**). These tantalizing results suggest a new
367 role of CDH1 in restricting the immune evasion of tumor cells beyond its canonical role
368 in cell-cell adhesion.
369
370 Previously, two distinct molecular subtypes of GAC were introduced:
371 mesenchymal phenotype (MP) and epithelial phenotype (EP) [49,50]. Since only the
372 DGAC1 subtype is linked with CDH1 downregulation and EMT (**Fig. 2A**), the DGAC1
373 subtype might belong to the MP subtype, which is associated with poor survival and
374 chemotherapy resistance [50]. Unlike DGAC1, DGAC2 does not show CDH1 loss and
375 EMT. Instead, DGAC2 is associated with RHOA activation (**Fig. 2D**), which might
376 explain how the DGAC2 subtype also exhibits diffuse-cell morphology without CDH1
377 loss. It should be noted that among several genetic mutations in GAC, including DGAC
378 and intestinal-type gastric cancer (IGC), the *CDH1* (20-30%) and *RHOA* (15-25%)
379 mutations are dominantly found in DGAC but IGC [38,51,52].
380
381 E-cadherin mediates cell-cell interaction via homophilic interaction with other E-
382 cadherin proteins from neighboring cells. The cytoplasmic domain of E-cadherin is
383 physically associated with Catenin proteins (α, β, γ, and p120) and actin cytoskeleton,
384 which plays a pivotal role in maintaining epithelial cell polarity and integrity [53].
385 Unexpectedly, scRNA-seq analyses of GOs showed that Cdh1 loss aberrantly alters
386 cell plasticity, cellular origin (from Aqp5+ to proliferating cells) (**Fig. 5M**), and cell
387 differentiation status (**Fig. 5N, O**) with distinct transcriptional signatures (**Fig. 5E-G**).
388 Furthermore, CDH1 loss activates EZH2 regulon and EZH2 blockade suppresses EKP
389 tumor growth (**Fig. 7**). Therefore, it should be determined whether EZH2-induced
390 transcriptional reprogramming mediates CDH1 loss-induced aberrant cell plasticity.
391
392 EZH2 modulates gene expression in various ways: gene repression via PRC2-
393 dependent histone methylation, PRC2-dependent non-histone protein methylation, or

394 gene activation via transcriptional activator complex. The detailed mechanisms of how
395 EZH2 is engaged in CDH1 loss-associated DGAC tumorigenesis remain to be
396 determined. Nonetheless, given that an EZH2 inhibitor (tazemetostat) is clinically
397 available, targeting EZH2 would be a viable option for the DGAC1 subtype in addition
398 to T cell-based ICIs. The use of epigenetic modulators has been found to enhance the
399 infiltration of effector T cells, suppress tumor progression, and improve the therapeutic
400 effectiveness of PD-L1 checkpoint blockade in prostate or head and neck cancer [54,55].
401 Additionally, pharmacological inhibition of EZH2 has been shown to inhibit tumor
402 growth and enhance the efficacy of anti-CTLA-4 treatment in bladder cancer [56]. Given
403 the enriched expression of immune checkpoints in DGAC1 (**Fig. 3H, M**), a combination
404 therapy involving EZH2 inhibitors and ICIs may hold potential benefits for DGAC1
405 patients.
406
407 The remaining question is how CDH1 loss activates the EZH2 regulon.
408 Mesenchymal cells re-wire PI3K/AKT signaling to stimulate cell proliferation [57].
409 Additionally, it was shown that PI3K/AKT signaling is required for EZH2 activity in
410 *KRAS*^{G12D} mutant cells [58]. Thus, it is plausible that EMT-activated PI3K/AKT signaling
411 might activate EZH2. Consistent with this, compared to DGAC2, the DGAC1 subtype
412 shows high scores for EMT and PI3K/AKT/MTOR pathways, and EZH2 downstream
413 target gene expression (**Fig. 2A, C, 7I, J**).
414
415 Limitations of scRNA-seq include relatively shallow sequencing depth and
416 restricted information not overcoming intra-tumoral heterogeneity. Thus, increasing the
417 number of scRNA-seq datasets and spatial transcriptomics should follow in future
418 studies. Furthermore, although this is the first stratification of DGAC by single-cell
419 transcriptome, the pathological relevance of CDH1 status (or alternative molecular
420 signatures; **Fig. 1L**) with ICI response remains to be clinically demonstrated.
421
422 Together, our study stratifies DGAC patients by integrative single-cell
423 transcriptomics with experimental validation and unravels an unexpected role of E-

cadherin in restricting transcriptional reprogramming and immune evasion of DGAC, which provides new insight into the biology of DGAC tumorigenesis and helps improve immunotherapy efficacy.

Author contributions

G.Z., Y.H., and J.-I.P. conceived and designed the experiments. G.Z., Y.H., S.Z., K.-P.K., K.-B.K., J.Z., and S.J. performed the experiments. G.Z., Y.H., S.Z., K.-P.K., K.-B.K., S.S., J.A.A., and J.-I.P. analyzed the data. M.P.P., Y.F., S.S., and J.A.A. provided the sequencing files and clinical data for human scRNA-seq analyses. N.N. and H.W. read and analyzed the stained slides. G.Z., Y.H., S.Z., K.-P.K., K.-B.K., and J.-I.P. wrote the manuscript.

Acknowledgments

We are grateful to Pierre D. McCrea, Malgorzata Kloc, Rachael Miller, and Adriana Paulucci for their insightful comments. This work was supported by the Cancer Prevention and Research Institute of Texas (RP200315 to J.-I.P.). The core facilities at MD Anderson (DNA Sequencing and Genetically Engineered Mouse Facility) were supported by National Cancer Institute Cancer Center Support Grant (P30 CA016672). This work was performed at the Single Cell Genomics Core at BCM partially supported by NIH-shared instrument grants (S10OD023469, S10OD025240) and P30EY002520.

Methods

Mice

All mouse experiments were approved by the MD Anderson Institutional Animal Care and Use Committee and performed under MD Anderson guidelines and the Association for Assessment and Accreditation of Laboratory Animal Care international standards. Compound transgenic mice *Kras*$^{LSL\text{-}G12D/+}$; *Trp53*$^{fl/fl}$ (KP) mice have been previously described [59]. C57BL/6 mice were purchased from the Jackson Laboratory (Maine, USA).

Gastric organoids generation

The protocol for generating gastric organoids (GOs) was previously described [60]. The mice were sacrificed, and the mouse stomach was collected, and the forestomach was removed. Then, the reserved stomach tissue was cut through the lesser curvature, and the stomach was rinsed with ice-cold PBS with 1% penicillin/streptomycin to remove blood. The tissue samples were carefully immersed in chelating buffer (sterile distilled water with 5.6 mmol/L Na_2HPO_4, 8.0 mmol/L KH_2PO_4, 96.2 mmol/L NaCl, 1.6 mmol/L KCl, 43.4 mmol/L sucrose, 54.9 mmol/L D-sorbitol, 0.5 mmol/L DL-dithiothreitol, pH 7) in a 10 cm dish, then the tissue was transferred to a dry dish. The epithelial layer was peeled and minced into pieces using forceps. Minced epithelial pieces were placed into 10 mL cold chelating buffer, followed by robust pipetting up and down to rinse the tissue until the supernatant was clear. A 20 mL chelating buffer was prepared with 10 mM EDTA under room temperature, and the tissue was incubated in there for 10 min. The tissue was tenderly pipetted gently once up and down, and the pieces were allowed to settle. The tissue was then moved to the clean bench. Most of the water was removed, and the tissue pieces were carefully placed in the middle of a sterile 10 cm dish. A glass microscopy slide was put on top of the tissue and pressure was added upon the slide until the tissue pieces seemed cloudy. The cloudy tissue pieces were then flushed from the slides in 30 mL of cold Advanced DMEM/F12. The large tissue fragments were allowed to sediment by gravity. The cloudy supernatant was transferred to two 15 ml tubes. The tubes were then centrifuged for 5 min at 200 g and 4°C. The supernatant was carefully removed and resuspended with Matrigel-medium mixture (12 µL Matrigel mix with 8 µL GOs culture medium/well). Approximately 40 glands per 20 µL Matrigel-medium mixture per well of a 48-well plate were seeded. The plate was steadily transferred to the incubator to let it solidify for 10 minutes. Then, 500 µL of GOs culture medium was added to cover the dome, and the plate was incubated at 37 °C with 5% CO2. The medium was changed every 2 days.

Gastric organoids culture

Table S1 was referred to for the culture medium ingredient. The organoids were passaged using the following steps: **1.** The culture medium was discarded. **2.** The Matrigel was scraped with a pipette tip and dissociated by pipetting. **3.** The organoids were collected from three wells (48-well) in the 15 mL tube with cold medium. **4.** The supernatant was discarded after centrifugation at 1000 RPM and 4°C. **5.** The

dissociated organoids were washed with 13 mL of cold 1´PBS, centrifuged (1000 RPM, 4 min), and the supernatant was removed. **6.** The organoids were resuspended in 1 mL of Trypsin-EDTA (0.05%). **7.** The sample was transferred to a 1.7 mL Eppendorf tube, then pipetted up and down. **8.** The sample was incubated in a 37 °C with 5% CO2 incubator for 30 min to 45 min. **9.** The tube was vibrated every 10 min. **10.** The organoid structure was further broken down by pipetting up and down. **11.** The sample was checked under the microscopy to ensure the organoids digested into cells. **12.** The sample was passed through the 35 μm cell strainer. **13.** The Trypsin was inactivated with 10% FBS medium and pipetted vigorously. **14.** The sample was collected in the 15 mL tube and centrifuged for 4 min at 1000 RPM. **15.** The supernatant was aspirated and the cells were resuspended with GOs culture medium. **16.** The cells were counted, viability was checked, and the appropriate number of cells was calculated. **17.** Every 8 μL of cell suspension was mixed with 12 μL of Matrigel as a mixture and seeded in the 48-well plate. **18.** The plate was transferred to the incubator and allowed to solidify for 10 minutes. **19.** 500 μL of GOs culture medium was added to cover the dome and incubated at 37 °C with 5% CO2. **20.** The medium was changed every 2 days.

The organoids were cryopreserved as follows: The organoids were dissociated following above **organoid passaging (step1-15)** protocol. The cells were then added with 10% volume of DMSO and transferred to the cryovials.

CRISPR/Cas9-based gene knockout in GOs

Knockout (KO) of *Cdh1* was performed by CRISPR/Cas9 genome editing using pLentiCRISPRv2 (Addgene plasmid #52961) according to Zhang laboratory's protocol [61]. Five single guide RNA (sgRNA) targeting *Cdh1* were designed using CRISPick (https://portals.broadinstitute.org/gppx/crispick/public) and cloned into a pLentiCRISPRv2-puro vector. An empty sgRNA vector was used as a negative control. The five targeting sequences against *Cdh1* were: #1: 5'-ATGAT GAAAA CGCCA ACGGG-3', #2: 5'-ACCCC CAAGT ACGTA CGCGG-3', #3: 5'-TTACC CTACA TACAC TCTGG-3', #4: 5'-AGGGA CAAGA GACCC CTCAA-3', and #5: 5'-CCCTC CAAAT CCGAT ACCTG-3'. sgRNA 1# (5'-ATGAT GAAAA CGCCA ACGGG-3') was successfully knock out *Cdh1* in GOs. See Table S2 for primer sequence to validate *Cdh1* knockout efficiency.

Lentivirus production and transduction

The HEK293T cells were co-transfected with 5 μg of constructs, 5 μg of plasmid Δ8.2 (Plasmid #8455, Addgene), and 3 μg of plasmid VSVG (Plasmid #8454, Addgene) in a 10 cm dish. The cells were incubated at 37°C, and the medium was replaced after 12 h. The virus-containing medium was collected 48 h after transfection. The organoids were dissociated following the **organoid passaging protocol (step 1-14)**, and the supernatant was aspirated, leaving the pellet. For transduction, 20 μL of cell suspension was used. The amount of polybrene (8 μg/mL) was calculated and mixed with virus-containing medium before adding to the cells. The polybrene containing virus medium was added to the cell pellet, and the cell suspension was transferred to a

1.7 mL Eppendorf Tube. The tube was centrifuged at 600 g at 37 °C for 1 h. Without disturbing the cell pellet, the tube was incubated in the 37 °C incubator for 4 h. The supernatant was then removed, and the cell pellet was resuspended with the required volume of GOs culture medium (8 μL for one well of 48-well plate) and placed on ice for cool down. The appropriate volume of pre-thawed Matrigel (12 μL for one well of 48-well plate) was added to the tube, and the dome was seeded in the middle of a 48-well plate. The plate was then incubated for 10 min at 37 °C with 5% CO2. GOs culture medium was added to the well. After 48 h, the infected organoids were selected with 2 μg/mL puromycin.

Adenovirus transduction

We used Adeno-Cre virus to treat *Kras*$^{LSL\text{-}G12D/+}$; *Trp53*$^{fl/fl}$ organoids. The protocol was previously described [62]. The cells were first dissociated from GOs as described in the **organoid passaging protocol (step 1-14)**. The cell number was counted, and the ratio of adenovirus: organoid cell was 1000 PFU/μL:1 cell. The cell suspension, virus-containing medium, and Matrigel were mixed, and the drop was placed in the center of the well. The cell suspension and virus-containing medium were mixed before adding GOs culture medium up to 8 μL. Then, 12 μL of Matrigel was added to the mixture on ice. The plate was incubated in the 37°C cell culture incubator for 15 min to allow the Matrigel to solidify. After 48 h, the infected organoids were treated with 10 μM Nutlin-3 to select Trp53 KO organoids. The primer sequence to validate *Trp53* KO and *Kras*$^{G12D/+}$ can be found in Table S2.

Organoid imaging and size measurement

After 7 days of organoid seeding in Matrigel, the size of the organoids was analyzed by measuring the volume under the microscope (ZEN software, ZEISS). To reduce the vulnerability of GOs, the measurements were conducted more than 3 passages after isolation from the knockout experiments. All experiments included more than 50 organoids per group.

Tissue microarray

DGAC cancer tissue microarray slides contained 114 patients' samples. Patients' information is shown in Table S4.

Histology and immunohistochemistry

All staining was performed as previously described [63]. For organoids staining, 7 days after seeding, GOs were collected by dissociating Matrigel mixture using ice-cold PBS and fixed in 4% paraformaldehyde at room temperature. For tumor tissue, excised tumors were washed with ice-cold PBS and fixed with formaldehyde at room temperature. After paraffin embedding, tumor tissue and organoid sections were mounted on microscope slides. For H&E staining, sections were incubated in hematoxylin for 3-5 min and eosin for 20-40 s. After washing with tap water, slides were dehydrated, and the coverslips were mounted with mounting media. For immunofluorescence staining, after blocking with 5% goat serum in PBS for 1 hr at room temperature, sections were incubated with primary antibodies (MKI67 [1:200],

CDH1 [1:200], CD3 [1:200], PDCD1 [1:200], TIM3 [1:200],) overnight at 4 °C and secondary antibody (1:250) for 1 hr at room temperature in dark. Sections were mounted with ProLong Gold antifade reagent with DAPI (Invitrogen). For immunohistochemistry staining, after blocking with 5% goat serum in PBS for 1 hr at room temperature, sections were incubated with primary antibodies (CDH1 [1:200], H3K27Me3 [1:200], H3K27Ac [1:200]) overnight at 4 °C and secondary antibody (1:250) for 1 hr at room temperature in dark. Incubate the slides in the DAB solution until tissue become brown and background still white. Observed under the microscope until the strongest signal shows and stop reaction with tap water wash. Used the same incubation time for same antibody on different slides. Sections were incubated in hematoxylin for 3-5 min and mounted with mounting media. Images were captured with the fluorescence microscope (Zeiss; AxioVision). See Table S3 for antibody information.

2D culture

The organoids were dissociated following the **organoid passaging protocol (step1-14)**. The supernatant was aspirated and then resuspended with DMEM + 10% FBS with 10 μM Y-27632, and the organoids were seeded on a 24-well plate. Cells were passaged every 3-5 days. After the third passage, Y-27632 was removed from the culture medium. DMEM supplemented with 10% FBS and 10% DMSO was used to freeze cells and store them in liquid nitrogen.

Allograft assay

Five-week-old C57BL/6 mice were maintained in the Division of Laboratory Animal Resources facility at MD Anderson. 2D-cultured KP and EKP cells (1×10^6) were injected subcutaneously into both flanks of mice. Tumor volume was calculated by measuring with calipers every 3-4 days (volume = (length $\times$ width^2)/2). Mice were euthanized, and tumors were collected at day 15. The excised tumors were photographed and paraffin-embedded for immunostaining. For GSK343 treatment, 2D-cultured EKP cells (1×10^6) were injected subcutaneously into both flanks of mice. After the tumors were palpable, we performed the first measurement with calipers. We divided the mice into two groups of three mice each and administered DMSO and GSK343 (20 mg/kg) intraperitoneally every other day. The initial tumor volumes between the two groups were comparable. Tumor volume was calculated by measuring with calipers every 3-4 days (volume = (length $\times$ width^2)/2). Mice were euthanized, and tumors were collected at day 20.

Cell proliferation assays

Cells (1×10^3) were seeded on a 60 mm dish, and the medium was replaced every 2 days. Cell proliferation was determined by crystal violet staining or Cell Counting Kit-8 (Dojindo Laboratories) according to the manufacturer's protocol. Plates were rinsed with 1× PBS, fixed with 4% paraformaldehyde solution for 20 min, and stained with crystal violet solution (0.1% crystal violet, 10% methanol) for 20 min, followed by rinsing with tap water.

Gastric organoids library preparation for scRNA-seq

For scRNA-seq, organoids from WT, KP, and EKP were collected 7 days after seeding and follow the **organoid passaging (step1-14)** protocol. After trypsin had been inactivated with 10% FBS DMEM, a single-cell suspension was collected by passing cells through a 70 μm cell strainer and followed by a 40 μm cell strainer. Each group was tagged with two CMO tags from the CellPlex kit (10× Genomics). The tagged cells of each group were pooled together with the same number of cells after being counted. Single cell Gene Expression Library was prepared according to Chromium Single Cell Gene Expression 3v3.1 kit with Feature Barcode technology for cell Multiplexing (10x Genomics). In Brief, tagged single cells, reverse transcription (RT) reagents, Gel Beads containing barcoded oligonucleotides, and oil were loaded on a Chromium controller (10x Genomics) to generate single cell GEMS (Gel Beads-In-Emulsions). Incubation of the GEM produced barcoded, full-length cDNA as well as barcoded DNA from the cell Multiplexing. Subsequently the GEMS are broken and pooled. Following cleanup using Dynabeads MyOne Silane Beads, full-length cDNA is amplified by PCR for library prep through fragmentation, end-repair, A-tailing, adaptor ligation and amplification, while the barcoded DNA from the cell Multiplexing is amplified for library prep via PCR to add sequencing primers. The cDNA library was sequenced on an Illumina NovaSeq platform (Novogene), mapped to the GRCm38/mm10 genome, and demultiplexed using CellRanger. The resulting count matrices files were analyzed in R (Seurat) or Python (Scanpy).

scRNA-seq - raw data processing, clustering, and annotation

We used Cell Ranger to perform demultiplexing and reads alignment of sequencing raw data for the scRNA-seq matrices generation. Ambient RNA and doublets were removed by SoupX [64] and Scrublet [65], respectively. Scanpy[66] was used for processing the scRNA-seq data. For the organoid dataset, cells with less than 50 genes expressed and more than 30% mitochondrial reads, 30% rpl reads, and 25% rps reads were removed. Genes expressed in less than 5 cells were removed. Then we normalized and log-transformed the gene expression for each cell. The percentages of mitochondrial reads, rpl reads, and rps reads were regressed before scaling the data. We reduced dimensionality and cluster the cells by Leiden (resolution=0.5). Cell lineages were annotated based on algorithmically defined marker gene expression for each cluster (sc.tl.rank_genes_groups, method='wilcoxon'). See Table S11, top 100 genes of each cluster were listed. For the DGAC dataset, cells with less than 100 genes expressed and more than 80% mitochondrial reads, 30% rpl reads, and 25% rps reads were removed. Genes expressed in less than 25 cells were removed. Normalization, log-transformation, regression, dimensionality reduction, and Leiden clustering (resolution=1) were the same as the way we use in organoids. Cell lineages were annotated based on algorithmically defined marker gene expression for each cluster (sc.tl.rank_genes_groups, method='t-test'). See Table S6, S7, and S8 for details, top 100 genes of each cluster or type were listed. For the DGAC dataset merged with normal stomach dataset, cells with less than 100 genes expressed and more than 100% mitochondrial reads, 40% rpl reads, and 30% rps reads were removed. Genes

expressed in less than 25 cells were removed. Normalization, log-transformation, regression, dimensionality reduction, and Leiden clustering (resolution=1) were the same as the way we use in organoids. Cell lineages were annotated based on algorithmically defined marker gene expression for each cluster (sc.tl.rank_genes_groups, method='t-test'). See Table S10 for details, top 100 genes of each cluster were listed. More information about the software and algorithms used in this study is shown in Table S13.

Cell lineage trajectory analysis

We use the CytoTRACE [42] kernel of CellRank [42] to predict a pseudotemporal ordering of cells from initial states to terminal states for the organoid dataset. Briefly, scRNA-seq matrices were pre-processed in the same way as Scanpy did until the step of log-transformation. Then, CytoTRACE kernel was called to compute the cytotrace pseudotime and cell fate trajectories (n_pcs=30, n_neighbors=10). GPCCA estimator was initiated and the scRNA-seq matrices was performed a Schur decomposition. Next, the terminal (backward=False, n_states=3 and initial (backward=True, n_states=1) macro-states were optimized based on the best eigenvalues with high confidence (>0.95), respectively. Finally, the CellRank corrected and cytotrace pseudotime directed PAGA [67] were generated.

Proportion difference analysis

The cell number of each cluster were retrieved by Scanpy (adata.obs['leiden'].value_counts()). We analyzed and plotted the differences between clusters from the two datasets using the GraphPad Prism 9.4. Then we grouped each cell cluster from the integrated dataset and compared the cluster differences between the two datasets.

Regulon analysis

For the gene regulatory network inference in organoids, we used the pySCENIC package [68] to compute the specific regulons for each cell cluster. The Loom file of each organoid dataset was used, and the regulon pattern-based UMAP was redrawn based on the AUCell scoring method [69]. Regulon specificity score (RSS) [70] and Z score were used to determine how specific the regulon is for one certain cell cluster. More specific the regulon is, the higher RSS or Z score is for one certain cluster. Following the criteria that RSS and Z score should be high at the same time, we identified 20 regulons that specific to EKP. These processes were repeated five times in each organoid dataset (WT, KP, and EKP).

Scissor analysis

To determine the pathology of murine organoids, we compared the transcriptomic similarity of the organoids scRNA-seq dataset and the bulk RNA-seq datasets of DGAC patients by Scissor package [41]. The RNA-seq data of tumor and the adjacent normal samples of DGAC patients were downloaded from the GDC data portal (TCGA-STAD). The murine genes were converted to human homologs by biomaRt. The Scissor analysis was performed by using the Cox regression model (alpha = 0.32).

714
715 **Cell-cell communication analysis**
716 'CellChat' [71] package in R (https://www.r-project.org) was used to analysis the ligand-
717 receptor interaction-based cell-cell communication in scRNA-seq datasets. The
718 integrated dataset was processed, clustered, and annotated using the scanpy package
719 [72] in python, then transformed into .rds files. Transformed datasets were analyzed by
720 CellChat with default parameters (p-value threshold = 0.05).
721
722 **Pathway score analysis**
723 Pathway score was analyzed by Scanpy [72] with the 'scanpy.tl.score_genes' function [66].
724 The analysis was performed with default parameters and the reference genes from the
725 gene ontology biological process or the Kyoto Encyclopedia of Genes and Genomes
726 database [73,74]. The gene list for the score analysis is shown in Table S9.
727
728 **Human scRNA-seq data analysis**
729 The scRNA-seq data set of 20 DGAC patients' samples (Patients information is shown
730 in Table S5) has been previous reported from our group and the detailed clinical and
731 histopathological characteristics are described (EGAS00001004443) [26]. The scRNA-seq
732 data set of the 29 normal adjacent stomachs (GSE150290) [34] was extracted from the
733 Gene Expression Omnibus (GEO) database and analyzed with Scanpy and Python [72].
734 The 20 DGAC patients' datasets were integrated and clustered by Scanpy [72] for the
735 subclassification of DGACs based on CDH1 inactivation. The 20 DGAC patients'
736 datasets and 29 normal adjacent stomachs were integrated and clustered in Scanpy [72]
737 for later infercnvpy analysis. "Harmony" [75] algorithm was used to remove batch effects.
738 Then, the dendrogram and correlation matrix heatmap were plotted with Scanpy [72]. The
739 dendrogram shows the distance of each dataset based on principal component
740 analysis, and the correlation matrix heatmap shows Pearson correlation by a color
741 spectrum.
742
743 **Copy number variation analysis**
744 To detect the genomic stability of groups DGAC1, DGAC2, and DGAC3, we performed
745 copy number variations (CNVs) inference from the gene expression data using the
746 Python package infercnvpy (https://icbi-lab.github.io/infercnvpy/index.html). We
747 performed infercnvpy on DGAC1, DGAC2, and DGAC3 using the Normal group (29
748 human normal adjacent stomachs) as reference. The gene ordering file which is
749 containing the chromosomal start and end position for each gene was created from the
750 human GRCh38 assembly. The GRCh38 genomic positions annotated file was
751 downloaded from https://support.10xgenomics.com/single-cell-gene-
752 expression/software/downloads/latest. Infercnvpy was used to plot chromosome
753 heatmap and CNV scores in the UMAP.
754
755 **Gene set enrichment analysis (GESA)**
756 GSEA was conducted via the R package "fgsea" [76] according to the DEG list generated
757 by Scanpy. The enrichment value was calculated and plotted with the fgsea package
758 (permutation number = 2,000).

Public sequencing database

All TCGA cancer patients' sequencing data referenced in this study were obtained from the TCGA database at cBioPortal Cancer Genomics (http://www.cbioportal.org).

Data availability

scRNA-seq data are available via the GEO database (GSE226266; log-in token for reviewers: ###).

Code availability

The code used to reproduce the analyses described in this manuscript can be accessed via GitHub (https://github.com/jaeilparklab/EKP_DGAC_project) and will also be available upon request.

Statistical analyses

GraphPad Prism 9.4 (Dogmatics) was used for statistical analyses. The Student's *t*-test was used to compare two samples. The one-way ANOVA was used to compare multiple samples. P values < 0.05 were considered statistically significant. Error bars indicate the standard deviation (s.d.) otherwise described in Figure legends.

References

1 Sung, H. *et al.* Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* **71**, 209-249 (2021). https://doi.org:10.3322/caac.21660

2 Iyer, P., Moslim, M., Farma, J. M. & Denlinger, C. S. Diffuse gastric cancer: histologic, molecular, and genetic basis of disease. *Transl Gastroenterol Hepatol* **5**, 52 (2020). https://doi.org:10.21037/tgh.2020.01.02

3 Chen, Y. C. *et al.* Clinicopathological Variation of Lauren Classification in Gastric Cancer. *Pathol Oncol Res* **22**, 197-202 (2016). https://doi.org:10.1007/s12253-015-9996-6

4 Korivi, B. R. *et al.* Intestinal and diffuse gastric cancer: a retrospective study comparing primary sites. *Clin Imaging* **56**, 33-40 (2019). https://doi.org:10.1016/j.clinimag.2019.03.002

5 Nie, R. C. *et al.* Clinicopathological Characteristics and Prognostic Value of Signet Ring Cells in Gastric Carcinoma: A Meta-Analysis. *J Cancer* **8**, 3396-3404 (2017). https://doi.org:10.7150/jca.21017

6 Sanjeevaiah, A., Cheedella, N., Hester, C. & Porembka, M. R. Gastric Cancer: Recent Molecular Classification Advances, Racial Disparity, and Management Implications. *J Oncol Pract* **14**, 217-224 (2018). https://doi.org:10.1200/JOP.17.00025

7 Henson, D. E., Dittus, C., Younes, M., Nguyen, H. & Albores-Saavedra, J. Differential trends in the intestinal and diffuse types of gastric carcinoma in the United States, 1973-2000: increase in the signet ring cell type. *Arch Pathol Lab Med* **128**, 765-770 (2004). https://doi.org:10.5858/2004-128-765-DTITIA

8 van der Kaaij, R. T. *et al.* A population-based study on intestinal and diffuse type adenocarcinoma of the oesophagus and stomach in the Netherlands between 1989 and 2015. *Eur J Cancer* **130**, 23-31 (2020). https://doi.org:10.1016/j.ejca.2020.02.017

9 Ajani, J. A. *et al.* Gastric Cancer, Version 2.2022, NCCN Clinical Practice Guidelines in Oncology. *J Natl Compr Canc Netw* **20**, 167-192 (2022). https://doi.org:10.6004/jnccn.2022.0008

10 Muro, K. *et al.* Pan-Asian adapted ESMO Clinical Practice Guidelines for the management of patients with metastatic gastric cancer: a JSMO-ESMO initiative endorsed by CSCO, KSMO, MOS, SSO and TOS. *Ann Oncol* **30**, 19-33 (2019). https://doi.org:10.1093/annonc/mdy502

11 Garcia-Pelaez, J., Barbosa-Matos, R., Gullo, I., Carneiro, F. & Oliveira, C. Histological and mutational profile of diffuse gastric cancer: current knowledge and future challenges. *Mol Oncol* **15**, 2841-2867 (2021). https://doi.org:10.1002/1878-0261.12948

12 Hironaka, S. *et al.* Randomized, open-label, phase III study comparing irinotecan with paclitaxel in patients with advanced gastric cancer without severe peritoneal metastasis after failure of prior combination chemotherapy using fluoropyrimidine plus platinum: WJOG 4007 trial. *J Clin Oncol* **31**, 4438-4444 (2013). https://doi.org:10.1200/JCO.2012.48.5805

13 Sym, S. J. *et al.* Salvage chemotherapy with biweekly irinotecan, plus 5-fluorouracil and leucovorin in patients with advanced gastric cancer previously treated with fluoropyrimidine, platinum, and taxane. *Am J Clin Oncol* **31**, 151-156 (2008). https://doi.org:10.1097/COC.0b013e31815878a2
14 Selim, J. H., Shaheen, S., Sheu, W. C. & Hsueh, C. T. Targeted and novel therapy in advanced gastric cancer. *Exp Hematol Oncol* **8**, 25 (2019). https://doi.org:10.1186/s40164-019-0149-6
15 Korfer, J., Lordick, F. & Hacker, U. T. Molecular Targets for Gastric Cancer Treatment and Future Perspectives from a Clinical and Translational Point of View. *Cancers (Basel)* **13** (2021). https://doi.org:10.3390/cancers13205216
16 Mazzarella, L. *et al.* The evolving landscape of 'next-generation' immune checkpoint inhibitors: A review. *Eur J Cancer* **117**, 14-31 (2019). https://doi.org:10.1016/j.ejca.2019.04.035
17 Boku, N. *et al.* Nivolumab in previously treated advanced gastric cancer (ATTRACTION-2): 3-year update and outcome of treatment beyond progression with nivolumab. *Gastric Cancer* **24**, 946-958 (2021). https://doi.org:10.1007/s10120-021-01173-w
18 Janjigian, Y. Y. *et al.* First-line nivolumab plus chemotherapy versus chemotherapy alone for advanced gastric, gastro-oesophageal junction, and oesophageal adenocarcinoma (CheckMate 649): a randomised, open-label, phase 3 trial. *Lancet* **398**, 27-40 (2021). https://doi.org:10.1016/S0140-6736(21)00797-2
19 Wainberg, Z. A. *et al.* Efficacy of Pembrolizumab Monotherapy for Advanced Gastric/Gastroesophageal Junction Cancer with Programmed Death Ligand 1 Combined Positive Score >/=10. *Clin Cancer Res* **27**, 1923-1931 (2021). https://doi.org:10.1158/1078-0432.CCR-20-2980
20 Teng, M. W., Ngiow, S. F., Ribas, A. & Smyth, M. J. Classifying Cancers Based on T-cell Infiltration and PD-L1. *Cancer Res* **75**, 2139-2145 (2015). https://doi.org:10.1158/0008-5472.CAN-15-0255
21 Ge, S. *et al.* A proteomic landscape of diffuse-type gastric cancer. *Nat Commun* **9**, 1012 (2018). https://doi.org:10.1038/s41467-018-03121-2
22 Cristescu, R. *et al.* Molecular analysis of gastric cancer identifies subtypes associated with distinct clinical outcomes. *Nat Med* **21**, 449-456 (2015). https://doi.org:10.1038/nm.3850
23 Lei, Z. *et al.* Identification of molecular subtypes of gastric cancer with different responses to PI3-kinase inhibitors and 5-fluorouracil. *Gastroenterology* **145**, 554-565 (2013). https://doi.org:10.1053/j.gastro.2013.05.010
24 Hansford, S. *et al.* Hereditary Diffuse Gastric Cancer Syndrome: CDH1 Mutations and Beyond. *JAMA Oncol* **1**, 23-32 (2015). https://doi.org:10.1001/jamaoncol.2014.168
25 van Roy, F. & Berx, G. The cell-cell adhesion molecule E-cadherin. *Cell Mol Life Sci* **65**, 3756-3788 (2008). https://doi.org:10.1007/s00018-008-8281-1

26 Wang, R. *et al.* Single-cell dissection of intratumoral heterogeneity and lineage diversity in metastatic gastric adenocarcinoma. *Nat Med* **27**, 141-151 (2021). https://doi.org:10.1038/s41591-020-1125-8
27 Gunes, A. *et al.* Thioredoxin interacting protein promotes invasion in hepatocellular carcinoma. *Oncotarget* **9**, 36849-36866 (2018). https://doi.org:10.18632/oncotarget.26402
28 Li, Y. *et al.* Hypoxia induced high expression of thioredoxin interacting protein (TXNIP) in non-small cell lung cancer and its prognostic effect. *Asian Pac J Cancer Prev* **16**, 2953-2958 (2015). https://doi.org:10.7314/apjcp.2015.16.7.2953
29 Gao, Y. *et al.* Decreased expression of TXNIP predicts poor prognosis in patients with clear cell renal cell carcinoma. *Oncol Lett* **19**, 763-770 (2020). https://doi.org:10.3892/ol.2019.11165
30 Mascarenhas, J. B. *et al.* EVL is a novel focal adhesion protein involved in the regulation of cytoskeletal dynamics and vascular permeability. *Pulm Circ* **11**, 20458940211049002 (2021). https://doi.org:10.1177/20458940211049002
31 Hu, L. D., Zou, H. F., Zhan, S. X. & Cao, K. M. EVL (Ena/VASP-like) expression is up-regulated in human breast cancer and its relative expression level is correlated with clinical stages. *Oncol Rep* **19**, 1015-1020 (2008).
32 Yang, H. *et al.* Stress-glucocorticoid-TSC22D3 axis compromises therapy-induced antitumor immunity. *Nat Med* **25**, 1428-1441 (2019). https://doi.org:10.1038/s41591-019-0566-4
33 Ooki, A. & Yamaguchi, K. The dawn of precision medicine in diffuse-type gastric cancer. *Ther Adv Med Oncol* **14**, 17588359221083049 (2022). https://doi.org:10.1177/17588359221083049
34 Kim, J. *et al.* Single-cell analysis of gastric pre-cancerous and cancer lesions reveals cell lineage diversity and intratumoral heterogeneity. *NPJ Precis Oncol* **6**, 9 (2022). https://doi.org:10.1038/s41698-022-00251-1
35 Jin, S. *et al.* Inference and analysis of cell-cell communication using CellChat. *Nat Commun* **12**, 1088 (2021). https://doi.org:10.1038/s41467-021-21246-9
36 Mimata, A., Fukamachi, H., Eishi, Y. & Yuasa, Y. Loss of E-cadherin in mouse gastric epithelial cells induces signet ring-like cells, a possible precursor lesion of diffuse gastric cancer. *Cancer Sci* **102**, 942-950 (2011). https://doi.org:10.1111/j.1349-7006.2011.01890.x
37 Hayakawa, Y. *et al.* Mist1 Expressing Gastric Stem Cells Maintain the Normal and Neoplastic Gastric Epithelium and Are Supported by a Perivascular Stem Cell Niche. *Cancer Cell* **28**, 800-814 (2015). https://doi.org:10.1016/j.ccell.2015.10.003
38 Cancer Genome Atlas Research, N. Comprehensive molecular characterization of gastric adenocarcinoma. *Nature* **513**, 202-209 (2014). https://doi.org:10.1038/nature13480
39 Korsunsky, I. *et al.* Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat Methods* **16**, 1289-1296 (2019). https://doi.org:10.1038/s41592-019-0619-0

40 Tan, S. H. *et al.* AQP5 enriches for stem cells and cancer origins in the distal stomach. *Nature* **578**, 437-443 (2020). https://doi.org:10.1038/s41586-020-1973-x

41 Sun, D. *et al.* Identifying phenotype-associated subpopulations by integrating bulk and single-cell sequencing data. *Nat Biotechnol* **40**, 527-538 (2022). https://doi.org:10.1038/s41587-021-01091-3

42 Gulati, G. S. *et al.* Single-cell transcriptional diversity is a hallmark of developmental potential. *Science* **367**, 405-411 (2020). https://doi.org:10.1126/science.aax0249

43 Kelley, D. R., Snoek, J. & Rinn, J. L. Basset: learning the regulatory code of the accessible genome with deep convolutional neural networks. *Genome Res* **26**, 990-999 (2016). https://doi.org:10.1101/gr.200535.115

44 Lee, M. G. *et al.* Demethylation of H3K27 regulates polycomb recruitment and H2A ubiquitination. *Science* **318**, 447-450 (2007). https://doi.org:10.1126/science.1149042

45 Verma, S. K. *et al.* Identification of Potent, Selective, Cell-Active Inhibitors of the Histone Lysine Methyltransferase EZH2. *ACS Med Chem Lett* **3**, 1091-1096 (2012). https://doi.org:10.1021/ml3003346

46 Tong, M. *et al.* Phosphoproteomics Enables Molecular Subtyping and Nomination of Kinase Candidates for Individual Patients of Diffuse-Type Gastric Cancer. *iScience* **22**, 44-57 (2019). https://doi.org:10.1016/j.isci.2019.11.003

47 Kim, S. K. *et al.* Identification of a molecular signature of prognostic subtypes in diffuse-type gastric cancer. *Gastric Cancer* **23**, 473-482 (2020). https://doi.org:10.1007/s10120-019-01029-4

48 Fukamachi, H. *et al.* A subset of diffuse-type gastric cancer is susceptible to mTOR inhibitors and checkpoint inhibitors. *J Exp Clin Cancer Res* **38**, 127 (2019). https://doi.org:10.1186/s13046-019-1121-3

49 Wang, R. *et al.* Multiplex profiling of peritoneal metastases from gastric adenocarcinoma identified novel targets and molecular subtypes that predict treatment response. *Gut* **69**, 18-31 (2020). https://doi.org:10.1136/gutjnl-2018-318070

50 Oh, S. C. *et al.* Clinical and genomic landscape of gastric cancer with a mesenchymal phenotype. *Nature Communications* **9**, 1777 (2018). https://doi.org:10.1038/s41467-018-04179-8

51 Kakiuchi, M. *et al.* Recurrent gain-of-function mutations of RHOA in diffuse-type gastric carcinoma. *Nat Genet* **46**, 583-587 (2014). https://doi.org:10.1038/ng.2984

52 Wang, K. *et al.* Whole-genome sequencing and comprehensive molecular profiling identify new driver mutations in gastric cancer. *Nat Genet* **46**, 573-582 (2014). https://doi.org:10.1038/ng.2983

53 McCrea, P. D. & Park, J. I. Developmental functions of the P120-catenin sub-family. *Biochim Biophys Acta* **1773**, 17-33 (2007). https://doi.org:10.1016/j.bbamcr.2006.06.009

54 Jadhav, R. R. *et al.* Epigenetic signature of PD-1+ TCF1+ CD8 T cells that act as resource cells during chronic viral infection and respond to PD-1 blockade. *Proc Natl Acad Sci U S A* **116**, 14113-14118 (2019). https://doi.org:10.1073/pnas.1903520116

55 Weber, E. W. *et al.* Transient rest restores functionality in exhausted CAR-T cells through epigenetic remodeling. *Science* **372** (2021). https://doi.org:10.1126/science.aba1786

56 Wherry, E. J. & Kurachi, M. Molecular and cellular insights into T cell exhaustion. *Nat Rev Immunol* **15**, 486-499 (2015). https://doi.org:10.1038/nri3862

57 Salt, M. B., Bandyopadhyay, S. & McCormick, F. Epithelial-to-mesenchymal transition rewires the molecular path to PI3K-dependent proliferation. *Cancer Discov* **4**, 186-199 (2014). https://doi.org:10.1158/2159-8290.CD-13-0520

58 Riquelme, E. *et al.* Modulation of EZH2 Expression by MEK-ERK or PI3K-AKT Signaling in Lung Cancer Is Dictated by Different KRAS Oncogene Mutations. *Cancer Res* **76**, 675-685 (2016). https://doi.org:10.1158/0008-5472.CAN-15-1141

59 Kim, M. J. *et al.* PAF remodels the DREAM complex to bypass cell quiescence and promote lung tumorigenesis. *Mol Cell* **81**, 1698-1714 e1696 (2021). https://doi.org:10.1016/j.molcel.2021.02.001

60 Bartfeld, S. *et al.* In vitro expansion of human gastric epithelial stem cells and their responses to bacterial infection. *Gastroenterology* **148**, 126-136 e126 (2015). https://doi.org:10.1053/j.gastro.2014.09.042

61 Ran, F. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nature protocols* **8**, 2281-2308 (2013).

62 Ko, K. P., Zhang, J. & Park, J. I. Establishing transgenic murine esophageal organoids. *STAR Protoc* **3**, 101317 (2022). https://doi.org:10.1016/j.xpro.2022.101317

63 Jung, Y. S. *et al.* Deregulation of CRAD-controlled cytoskeleton initiates mucinous colorectal cancer via beta-catenin. *Nat Cell Biol* **20**, 1303-1314 (2018). https://doi.org:10.1038/s41556-018-0215-z

64 Young, M. D. & Behjati, S. SoupX removes ambient RNA contamination from droplet-based single-cell RNA sequencing data. *Gigascience* **9** (2020). https://doi.org:10.1093/gigascience/giaa151

65 Wolock, S. L., Lopez, R. & Klein, A. M. Scrublet: Computational Identification of Cell Doublets in Single-Cell Transcriptomic Data. *Cell Syst* **8**, 281-291 e289 (2019). https://doi.org:10.1016/j.cels.2018.11.005

66 Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biology* **19**, 15 (2018). https://doi.org:10.1186/s13059-017-1382-0

67 Wolf, F. A. *et al.* PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. *Genome Biol* **20**, 59 (2019). https://doi.org:10.1186/s13059-019-1663-x

68 Van de Sande, B. *et al.* A scalable SCENIC workflow for single-cell gene regulatory network analysis. *Nat Protoc* **15**, 2247-2276 (2020). https://doi.org:10.1038/s41596-020-0336-2

69 *AUCell: Identifying cells with active gene sets*, <https://bioconductor.org/packages/release/bioc/vignettes/AUCell/inst/doc/AUCell.html> (2022).

70 Suo, S. *et al.* Revealing the Critical Regulators of Cell Identity in the Mouse Cell Atlas. *Cell Rep* **25**, 1436-1445 e1433 (2018). https://doi.org:10.1016/j.celrep.2018.10.045

71 Jin, S. *et al.* Inference and analysis of cell-cell communication using CellChat. *Nature communications* **12**, 1-20 (2021).

72 Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol* **19**, 15 (2018). https://doi.org:10.1186/s13059-017-1382-0

73 Kanehisa, M. Toward pathway engineering: a new database of genetic and molecular pathways. *Sci. Technol. Jap.* **59**, 34-38 (1996).

74 Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**, 25-29 (2000). https://doi.org:10.1038/75556

75 Korsunsky, I. *et al.* Fast, sensitive and accurate integration of single-cell data with Harmony. *Nature methods* **16**, 1289-1296 (2019).

76 Korotkevich, G. *et al.* Fast gene set enrichment analysis. *BioRxiv*, 060012 (2021).

Figure Legends

Figure 1. CDH1 inactivation in DGAC patient tumors

A. Genetic alteration of the CDH1 based on the cBioPortal stomach cancer datasets (http://www.cbioportal.org). DGAC, diffuse-type gastric adenocarcinoma; SRCC, signet ring cell carcinoma; TAC, tubular adenocarcinoma; STAD, stomach adenocarcinoma; MAC, mucinous adenocarcinoma; PAC, papillary adenocarcinoma.

B, C. IHC staining of CDH1 in 114 DGAC patient tumor samples. The representative images are shown (B). Quantification of H score of CDH1 expression (C). *P* values were calculated using the one-way ANOVA; error bars: standard deviation (SD). Clinical information of 114 DGAC patients was showed in Table S4**.**

D. Merged batch-based integrated UMAPs of 20 DGAC patients; integration package: Harmony. Clinical information of 20 DGAC patients was showed in Table S5**.**

E. Merged Leiden-based integrated UMAP of 20 DGAC patients. Dashed line circle: epithelial cells. Epi: Epithelial cells; Myeloid: myeloid cells; Effector T: effector T cells; Naïve T: Naïve T cells; Exhausted T: Exhausted T cells.

F. Merged cell type-based UMAP of 20 DGAC patients. All cells were re-clustered according to the Leiden clusters and gathered as mega clusters. Dashed line circle: epithelial cells.

G. Epithelial cells were clustered by Leiden.

H. Correlation matrix plot of epithelial cells showing pair-wise correlations among all samples above. The dendrogram shows the distance of each dataset based on principal component analysis, and the Pearson correlation is displayed with a color spectrum. Groups of patients were categorized by dendrogram and correlation.

I. Merged and separated UMAPs of DGAC1 and DGAC2.

J. Feature plots of epithelial cells displaying *CDH1* expression.

K. Dot plots of epithelial cells of *CDH1* expression in different DGAC groups and individual patients.

L. Molecular signatures of DGAC1 and DGAC2 patients. Dot plots of epithelial cells of each gene in different subtypes and individual patient.

Figure 2. Molecular characterization of DGAC subtypes

A-H. Dot plots of EMT (A), FGFR2 (B), PI3K_AKT_MTOR (C), RHOA (D), MAPK (E), HIPPO (F), WNT (G), and TGFBETA (H) scores in two DGAC types. *P* values

were calculated by using a *t*-test. The genes included in each score are listed in Table S9.

I. Merged batch-based UMAP of 29 adjacent normal stomach tissue (Normal tissue) and 20 DGAC patients. Total cell numbers are 90455. Integration package: Harmony.

J. Merged Leiden-based integrated UMAPs of 29 adjacent normal stomach tissue (Normal tissue) and 20 DGAC patients. Epi: Epithelial cells; Myeloid: myeloid cells; Effector T: effector T cells; Naïve T: Naïve T cells; Exhausted T: Exhausted T cells. Top 100 genes of each cluster were showed in Table S10.

K. Merged cell type-based UMAP of 29 Normal tissue and 20 DGAC patients. All cells were re-clustered according to the Leiden clusters and gathered as mega clusters. Dashed line-circle: epithelial cells.

L. Separated UMAPs of Normal tissue and two types of DGACs. Dashed line-circle: epithelial cells.

M. CNV scores projected into the UMAP of the scRNA-seq dataset from adjacent normal stomach tissue, DGAC1, and DGAC2. Red: copy number gain (CNG); blue: copy number loss (CNL).

N. Leiden-based CNV plot showing the distribution of genomic alterations (gains and loss) in DGAC1 and DGAC2 compared with adjacent normal stomach tissue (Normal). Dark blue: CNV score low; yellow: CNV score high.

O. Statistics analysis of CNV score among Normal, DGAC1, and DGAC2. *P* values were calculated using the one-way ANOVA; error bars: SD.

Figure 3. Comparative analyses of immune landscapes of DGAC subtypes

A-B. Leiden-based and cell type-based UMAPs of DGAC1 and DGAC2.

C. Absolute and relative cell proportions of individual patients and DGAC subtypes. Patients list was ranked by the DGAC group that they belong.

D. Total cell-cell interactions (upper) and interaction strength (lower) from DGAC1 and DGAC2 were analyzed by using the CellChat package. More interactions were found in DGAC1.

E-F. Differential number of interactions between DGAC1 and DGAC2 using circle plots (E) and heatmap (F). Red (or blue) colored edges (E) and squares (F) represent increased (or decreased) signaling in the DGAC1 compared to DGAC2. The interaction between fibroblast and epithelial cells, and endothelial cells were decreased in DGAC1 compared to DGAC2, while the interaction of other cell types were increased.

G-H. Dot plots of exhausted T cell score (markers are included in that score: *LAG3*, *TIGIT*, *CTLA4*, and *HAVCR2*) and immune checkpoint score (markers are

included in that score: *CTLA4*, *PDCD1*, *PDCD1LG2*, and *CD274*). Genes that included in score analysis were showed in Table S9.

I. Cell type-based UMAP of 20 DGAC patients.

J-L. Feature plots of exhausted T cell score and immune checkpoint score in DGAC1 and DGAC2. *P* values were calculated using the Student's *t*-test; error bars: SD.

K, M. Dot plot of exhausted T cell score-related (K) and immune checkpoint (M)-related marker genes.

Figure 4. Establishment of genetically engineered gastric organoids with CDH1-inactivation

A. Genetic alteration of the *KRAS*, and *TP53* genes based on the cBioportal.

B. Illustration of the workflow for stomach tissue collection and dissociation, gene manipulation of the gastric organoids (GOs), GOs culture, and representative image of GOs. Three GO lines were generated, including WT, KP, and EKP. WT mice and KP mice were sacrificed to collect stomach tissue. After removing forestomach, stomach tissue was dissociated into single cell and culture as organoids. Adeno-Cre virus was used to treat $Kras^{LSL\text{-}G12D}$; $Trp53^{fl/fl}$ organoids to generate KP organoids, followed by nutlin-3 selection. After selection, EKP organoids were generated using CRISPR-mediated *Cdh1* KO from KP GOs.

C. Representative images of WT, KP, and EKP GOs at passage day 8. Scale bar: 200 μm.

D. Growth analysis for WT, KP, and EKP GOs in two passages at day 8 of each passage. P values were calculated using the one-way ANOVA; error bars: SD. ns: non-significant; **: $P < 0.01$; ***: $P < 0.001$. Numbers below each label represent the number of organoids.

E. Hematoxylin and eosin (H & E) staining of WT, KP, and EKP GOs.

F. MKI67 staining of WT, KP, and EKP GOs (n=5).

G. CDH1 staining of WT, KP, and EKP GOs.

H. Statistics analysis of MKI67 staining (Figure 4F). *P* values were calculated using the one-way ANOVA; error bars: SD. The representative images are shown.

Figure 5. scRNA-seq-based comparative analyses of genetically engineered GOs

A. Batch-based UMAPs of WT, KP, and EKP GOs. The Harmony integration package was used to remove the batch effect.

B. Leiden-based clustering UMAPs of WT, KP, and EKP GOs. Cell clusters were named by the top expressed genes.

C. Cell proportion analysis of WT, KP, and EKP GOs. Each color represents a different cell type. The color code is based on the cell types shown in Figure 5B.

D-G. Dot plot of marker genes (D), mucinous markers (E), and gastric epithelium stemness markers (F), and human DGAC-related diagnostic markers (G) in each cluster of WT, KP, and EKP GOs. *Krt7*, *Muc1*, and *Sox4* were enriched in EKP GOs.

H. Batch-based and Scissor-based UMAP of WT and EKP GOs generated by Scissor package. TCGA datasets of normal stomach and DGAC patients were utilized.

I-L. Feature plots of significant up or down regulated markers (*Muc1*, *Krt7*, *Mki67*, and *Aqp5*) from Figure 5E-G.

M. CytoTRACE–based cell lineage trajectory analysis of scRNA-seq datasets (WT, KP, and EKP GOs). Cells were clustered using the "Leiden" algorithm, the CytoTRACE and Scanpy packages (n_neighbors = 15, n_pcs = 50).

N. CytoTRACE pseudotime analysis of WT, KP, and EKP GOs. Cells-of-origin clusters were marked with larger dots in the lower panel.

O. PAGA analysis of WT, KP, and EKP GOs was performed and visualized with CytoTRACE package. Cells-of-origin clusters were marked with red circle. Arrows represent the differentiation trajectory.

P. Feature plots of *Hmgb2* and *Pclaf*. *P* values were calculated by using Wilcoxon rank-sum.

Figure 6. *CDH1* KO promotes KP-driven gastric tumorigenesis

A. Illustration of the workflow for 2D culture and subcutaneous transplantation.

B. Bright-field images of KP and EKP cells in low and high magnification.

C. Crystal violet staining of KP and EKP GOs-derived cells.

D. Bright-field images of KP and EKP allograft tumors; tumor incidence of allograft tumors.

E, F. Plot for tumor mass (E) and tumor size (F) assessment of KP and EKP allografts.

G. H & E staining of KP and EKP allograft tumors.

H-L. MKi67, E-Cadherin, CD3, PDCD1, and TIM3 staining of KP and EKP allograft tumors ($n \geq 3$). Left images: low magnification. Right images: high magnification. Scale bars were shown on the representative images.

M-P. Statistics analysis of MKi67, CD3, PDCD1, and TIM3 staining in Figure 6H, J-L. *P* values were calculated using Student's *t*-test; error bars: SD.

1169 **Figure 7. *CDH1* KO-activated EZH2 promotes gastric tumorigenesis**

1170 **A.** Integrated batch-based and regulon pattern-based UMAP for WT, KP, and EKP
1171 GOs. Six transcriptional modules were identified.

1172 **B.** Separated regulon patterns based UMAP for WT, KP, and EKP GOs.

1173 **C.** Flow chart of regulons selection process.

1174 **D.** Regulons enriched in WT, KP, and EKP GOs, based on Z Score. 32 regulons
1175 were highly expressed in EKP samples compared to WT and KP.

1176 **E.** Regulons enriched in WT, KP, and EKP GOs, based on Regulon Specificity
1177 Score (RSS). The top 20 were selected by Z score. The whole regulon list based
1178 on RSS was showed in Table S12.

1179 **F.** Venn diagram for the regulons from figure 7D and 7E. 20 regulons were
1180 overlapped.

1181 **G.** Dot plot of the regulons (WT, KP and EKP GOs) increased in TCGA DGAC
1182 patients.

1183 **H.** Regulon activity-based UMAP of Ezh2 in WT, KP, and EKP GOs. The cells with
1184 lighter color represent regulated by Ezh2.

1185 **I-J.** Dot plot and feature plots of EZH2 downstream target genes scores in the
1186 epithelial cells of DGAC1 and DGAC2. Gene list of EZH2 targeted genes was
1187 listed in Table S9.

1188 **K.** The level of H3K27Ac and H3K27Me3 expression in KP and EKP allografts.
1189 Quantification was displayed.

1190 **L.** Crystal violet staining of KP and EKP cells after GSK343 (EZH2 inhibitor, 96 hrs).

1191 **M-O.** Allograft transplantation of EKP cells followed by EZH2 inhibition. Bright-field
1192 images of EKP allograft tumors treated with DMSO and GSK343 (20 mg/kg)
1193 separately (M). Tumor mass of EKP allografts treated with DMSO and GSK343
1194 (20 mg/kg) after mice scarification (N). Tumor growth curve of EKP allografts
1195 treated with DMSO and GSK343 (20 mg/kg) after cell subcutaneous
1196 transplantation (O).

1197 *P* values were calculated using Student's *t*-test; error bars: SD.

1198

1199

1200

Supplementary Figures

Supplementary Figure S1. The way of clustering scRNA-seq datasets of 20 DGAC patients

A. Dot plots of epithelial cell, myeloid cell, B cell, plasma cell, T cell, effector T cell, naïve T cell, exhausted T cell, fibroblast, and endothelial cell markers in merged 20 DGAC patients scRNA-seq data.

B. Leiden-based heatmap of all cells of merged datasets with annotation in 20 DGAC patients. Top 100 genes of each cluster were showed in Table S6.

C. Leiden-based heatmap of epithelial cells of merged datasets in 20 DGAC patients. Top 100 genes of each cluster were showed in Table S7.

D. Type-based heatmap of epithelial cells of merged datasets in 20 DGAC patients. Top 100 genes of each type were showed in Table S8.

Supplementary Figure S2. DGAC specific pathway scores in DGAC subtypes.

A-H. Violin plots and feature plots of EMT, FGFR2, PI3K_AKT_MTOR, RHOA, MAPK, HIPPO, WNT, and TGFBETA score in DGAC1 and DGAC2. The genes that are included in each score are listed in Table S9. *P* values of each pathway between DGAC1 and DGAC2 were showed in Figure 2A-2H with dot plot.

Supplementary Figure S3. scRNA-seq analysis of 20 DGAC patients and 29 adjacent normal stomach tissue

A. Dot plots of epithelial cell, myeloid cell, B cell, plasma cell, T cell, effector T cell, naïve T cell, exhausted T cell, fibroblast, and endothelial cell markers in merged 20 DGAC patients and 29 adjacent normal stomach tissue scRNA-seq data.

B. Annotated Leiden-based integrated UMAPs of 20 DGAC patients and 29 adjacent normal stomach tissue. Epi: Epithelial cells; Myeloid: myeloid cells; Effector T: effector T cells; Naïve T: Naïve T cells; Exhausted T: Exhausted T cells; Endothelial: Endothelial cells.

C. Type-based heatmap of all cells of merged datasets in 20 DGAC patients and 29 adjacent normal stomach tissue.

Supplementary Figure S4. Individual cell type-based UMAP of each DGAC patient

A-B. Individual cell type-based UMAP of the patients in DGAC1 and DGAC2. DGAC1 patients were enriched with stromal cells, mainly T cells. DGAC2 patients were enriched with epithelial cells.

Supplementary Figure S5. fGSEA analysis of DGAC1 compared with DGAC2 based on GOBP, REACTOME, and WP datasets.

A-C. fGSEA analysis of DGAC1 compared with DGAC2 based on GOBP (**A**), REACTOME (**B**), and WP (**C**) datasets. GOBP: Gene ontology biological process; REACTOME: Reactome gene sets; WP: WikiPathways gene sets. Pathways related with immune response were enriched in DGAC1 based on GOBP and WP.

Supplementary Figure S6. fGSEA analysis of DGAC1 compared with DGAC2 based on BIOCARTA, PID, and KEGG datasets.

A-C. fGSEA analysis of DGAC1 compared with DGAC2 based on BIOCARTA (**A**), PID (**B**), and KEGG (**C**) datasets. BIOCARTA: BioCarta gene set; PID: PID gene sets; KEGG: KEGG gene sets. Pathways related with immune response were enriched in DGAC1 based on all three datasets.

Supplementary Figure S7. Comparative analyses of the expression of macrophage polarization and myeloid-derived suppressor cell markers of Normal tissue and DGAC patients

A-B. Dot plot of macrophage polymerization markers in DGAC1 and DGAC2. No significant difference of M1 markers between DGACs. For M2 markers, except for VEGFA was enriched in DGAC1, there were no significant difference of other M2 markers between DGACs.

C-D. Dot plot (**C**) and violin plot (**D**) of myeloid-derived suppressor cell (MDSC) score in DGAC1 and DGAC2. DGAC1 has higher MDSC score. *P* values were calculated by using a *t*-test. Gene list for calculate MDSC score was showed in Table S9.

Supplementary Figure S8. Validation of genetic engineering

Genotyping results of KP organoids. After adeno-Cre treatment, KP organoids lost *Trp53*, while KrasG12D was activated in KP organoids. After *Cdh1* CRISPR knock out (KO), we performed sanger sequencing to compare the sequence of *Cdh1* in WT and EKP. The five targeting sequences against *Cdh1* were showed in methods '**CRISPR/Cas9-based gene knockout in GOs**'. The primers used for genotyping were showed in Table S2.

Supplementary Figure S9. scRNA-seq analysis of mouse GOs

A. Illustration of the workflow for stomach tissue collection and dissociation, gene manipulation of the gastric organoids (GOs), sample preparation of multiplex scRNA sequencing.

B. Workflow of single cell library preparation.

C. Heatmap of each cell clusters of merged datasets, including WT, KP, and EKP.

D-F. Separate heatmap of each cell clusters of WT, KP, and EKP datasets, respectively.

Supplementary Figure S10. Feature plots of mucinous, stemness, and diagnostic markers in WT, KP, and EKP GOs

A-C. Feature plots of mucinous markers, gastric epithelium stemness markers, and DGAC-related diagnostic markers in WT, KP, and EKP organoids.

D. *P* values of the feature plots from figure S10A-S10C. *P* values were calculated by using Wilcoxon rank-sum. Red marked *P* values were significant ones (less than 0.05).

Supplementary Figure S11. EKP-specific regulons expression in the TCGA DGAC dataset and regulon activity-based UMAPs

A. The expression of 20 regulons in TCGA DGAC patients and normal stomach.

B. Regulon activity based UMAP of Gtf2b, Pole4, and Sox4. *P* values were calculated by using the Student's *t*-test; error bars: SD.

Figure 1

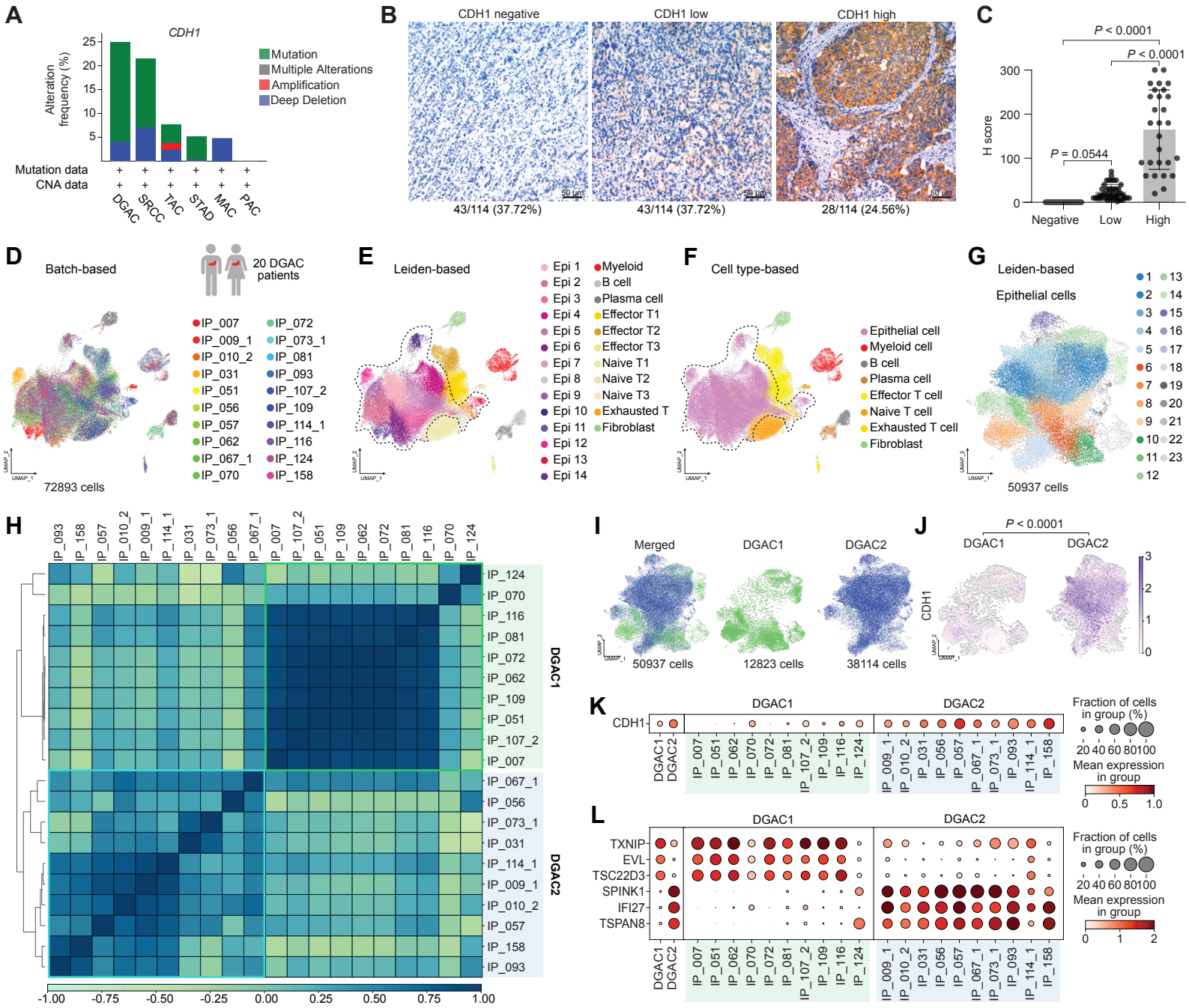

A
CDH1
Alteration frequency (%)
Mutation
Multiple Alterations
Amplification
Deep Deletion
Mutation data
CNA data
DGAC
SRCC
TAC
STAD
MAC
PAC
B
CDH1 negative
CDH1 low
CDH1 high
43/114 (37.72%)
43/114 (37.72%)
28/114 (24.56%)
C
P < 0.0001
P < 0.0001
P = 0.0544
H score
Negative
Low
High
D
Batch-based
20 DGAC patients
72893 cells
E
Leiden-based
F
Cell type-based
G
Leiden-based
Epithelial cells
50937 cells
H
DGAC1
DGAC2
I
Merged
DGAC1
DGAC2
50937 cells
12823 cells
38114 cells
J
P < 0.0001
DGAC1
DGAC2
CDH1
K
DGAC1
DGAC2
CDH1
Fraction of cells in group (%)
Mean expression in group
L
DGAC1
DGAC2
TXNIP
EVL
TSC22D3
SPINK1
IFI27
TSPAN8
Fraction of cells in group (%)
Mean expression in group

Figure 2

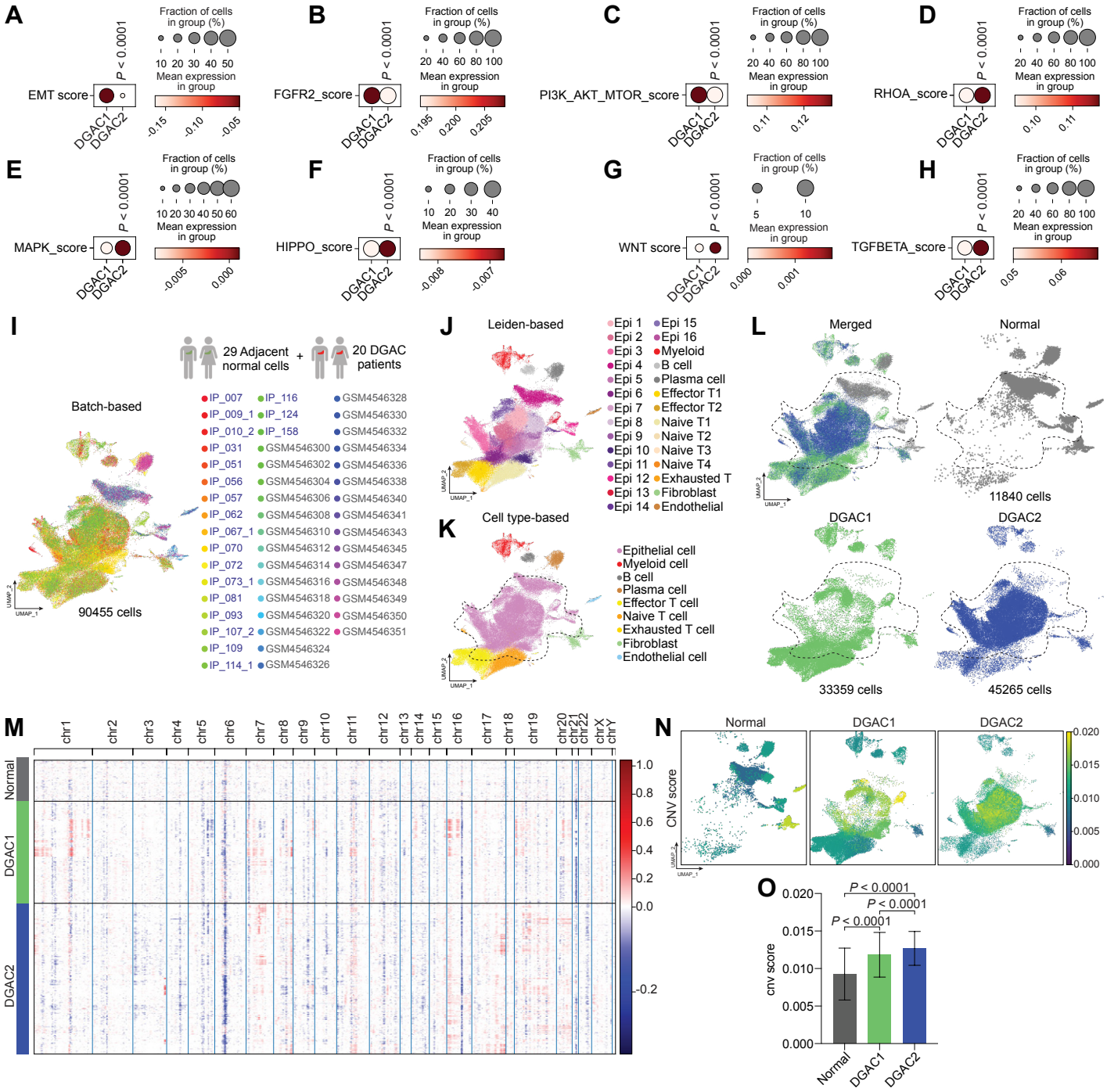

A
EMT score
P < 0.0001
Fraction of cells in group (%)
10 20 30 40 50
Mean expression in group
-0.15 -0.10 -0.05
DGAC1 DGAC2
B
FGFR2_score
P < 0.0001
Fraction of cells in group (%)
20 40 60 80 100
Mean expression in group
0.195 0.200 0.205
C
PI3K_AKT_MTOR_score
P < 0.0001
Fraction of cells in group (%)
20 40 60 80 100
Mean expression in group
0.11 0.12
D
RHOA_score
P < 0.0001
Fraction of cells in group (%)
20 40 60 80 100
Mean expression in group
0.10 0.11
E
MAPK_score
P < 0.0001
Fraction of cells in group (%)
10 20 30 40 50 60
Mean expression in group
-0.005 0.000
F
HIPPO_score
P < 0.0001
Fraction of cells in group (%)
10 20 30 40
Mean expression in group
-0.008 -0.007
G
WNT score
P < 0.0001
Fraction of cells in group (%)
5 10
Mean expression in group
0.000 0.001
H
TGFBETA_score
P < 0.0001
Fraction of cells in group (%)
20 40 60 80 100
Mean expression in group
0.05 0.06
I
29 Adjacent normal cells + 20 DGAC patients
Batch-based
90455 cells
IP_007 IP_116 GSM4546328
IP_009_1 IP_124 GSM4546330
IP_010_2 IP_158 GSM4546332
IP_031 GSM4546300 GSM4546334
IP_051 GSM4546302 GSM4546336
IP_056 GSM4546304 GSM4546338
IP_057 GSM4546306 GSM4546340
IP_062 GSM4546308 GSM4546341
IP_067_1 GSM4546310 GSM4546343
IP_070 GSM4546312 GSM4546345
IP_072 GSM4546314 GSM4546347
IP_073_1 GSM4546316 GSM4546348
IP_081 GSM4546318 GSM4546349
IP_093 GSM4546320 GSM4546350
IP_107_2 GSM4546322 GSM4546351
IP_109 GSM4546324
IP_114_1 GSM4546326
J
Leiden-based
Epi 1 Epi 15
Epi 2 Epi 16
Epi 3 Myeloid
Epi 4 B cell
Epi 5 Plasma cell
Epi 6 Effector T1
Epi 7 Effector T2
Epi 8 Naive T1
Epi 9 Naive T2
Epi 10 Naive T3
Epi 11 Naive T4
Epi 12 Exhausted T
Epi 13 Fibroblast
Epi 14 Endothelial
K
Cell type-based
Epithelial cell
Myeloid cell
B cell
Plasma cell
Effector T cell
Naive T cell
Exhausted T cell
Fibroblast
Endothelial cell
L
Merged
Normal
11840 cells
DGAC1
DGAC2
33359 cells
45265 cells
M
chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr13 chr14 chr15 chr16 chr17 chr18 chr19 chr20 chr21 chr22 chrX chrY
Normal
DGAC1
DGAC2
1.0 0.8 0.6 0.4 0.2 0.0 -0.2
N
Normal
DGAC1
DGAC2
CNV score
0.020 0.015 0.010 0.005 0.000
O
P < 0.0001
P < 0.0001
P < 0.0001
cnv score
0.020 0.015 0.010 0.005 0.000
Normal
DGAC1
DGAC2

Figure 3

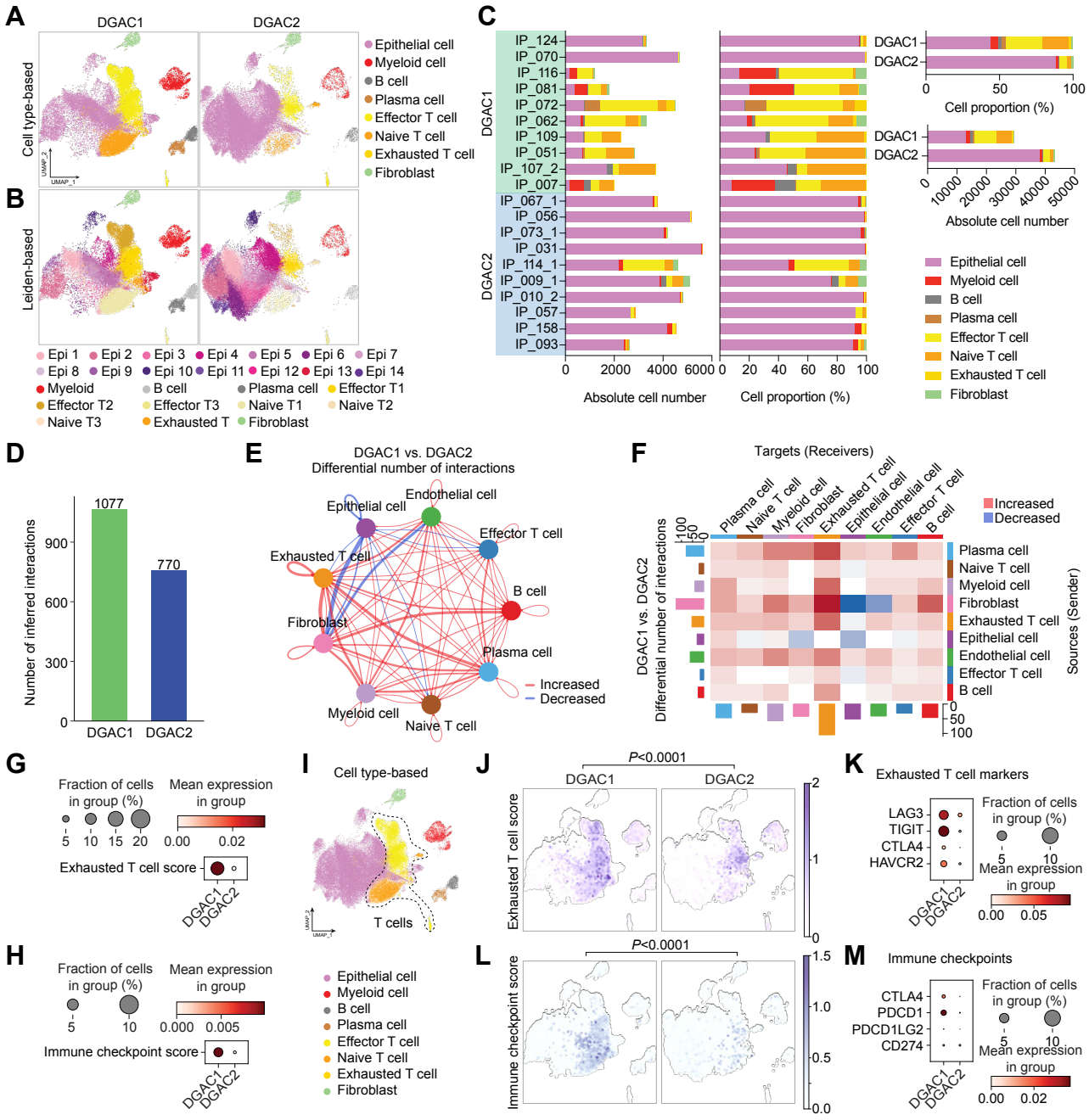

A
DGAC1
DGAC2
Cell type-based
Epithelial cell
Myeloid cell
B cell
Plasma cell
Effector T cell
Naive T cell
Exhausted T cell
Fibroblast
B
Leiden-based
Epi 1 Epi 2 Epi 3 Epi 4 Epi 5 Epi 6 Epi 7
Epi 8 Epi 9 Epi 10 Epi 11 Epi 12 Epi 13 Epi 14
Myeloid B cell Plasma cell Effector T1
Effector T2 Effector T3 Naive T1 Naive T2
Naive T3 Exhausted T Fibroblast
C
DGAC1
DGAC2
Absolute cell number
Cell proportion (%)
D
Number of inferred interactions
1077
770
DGAC1 DGAC2
E
DGAC1 vs. DGAC2
Differential number of interactions
Endothelial cell
Epithelial cell
Effector T cell
Exhausted T cell
B cell
Fibroblast
Plasma cell
Myeloid cell
Naive T cell
Increased
Decreased
F
Targets (Receivers)
Plasma cell
Naive T cell
Myeloid cell
Fibroblast
Exhausted T cell
Epithelial cell
Endothelial cell
Effector T cell
B cell
Increased
Decreased
DGAC1 vs. DGAC2
Differential number of interactions
Sources (Sender)
G
Fraction of cells in group (%)
Mean expression in group
Exhausted T cell score
DGAC1
DGAC2
I
Cell type-based
T cells
J
P<0.0001
DGAC1
DGAC2
Exhausted T cell score
K
Exhausted T cell markers
LAG3
TIGIT
CTLA4
HAVCR2
Fraction of cells in group (%)
Mean expression in group
H
Immune checkpoint score
L
P<0.0001
Immune checkpoint score
M
Immune checkpoints
CTLA4
PDCD1
PDCD1LG2
CD274

Figure 4

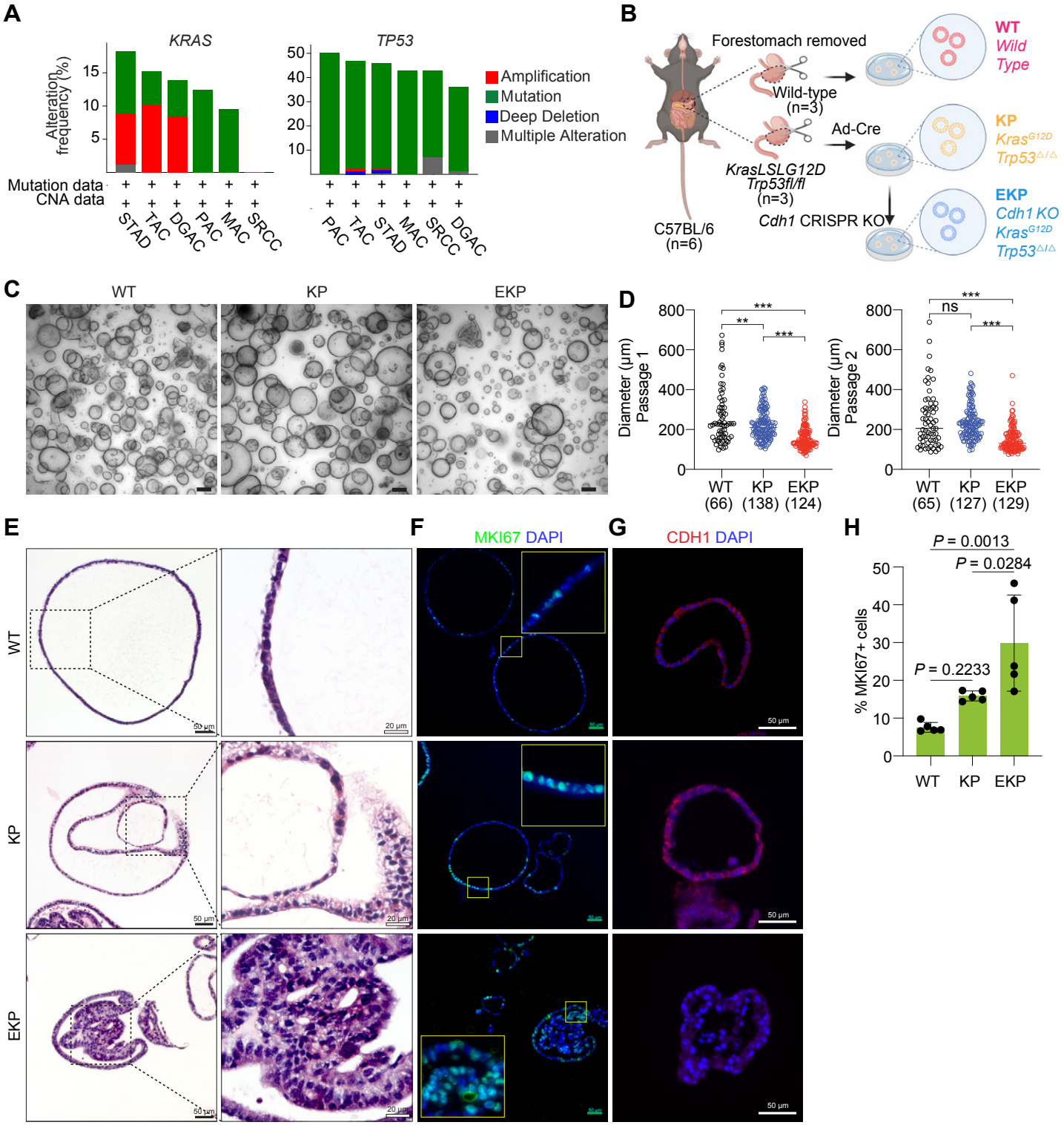

A
KRAS
TP53
Alteration frequency (%)
Amplification
Mutation
Deep Deletion
Multiple Alteration
Mutation data
CNA data
STAD
TAC
DGAC
PAC
MAC
SRCC
B
Forestomach removed
Wild-type (n=3)
Ad-Cre
KrasLSLG12D Trp53fl/fl (n=3)
Cdh1 CRISPR KO
C57BL/6 (n=6)
WT Wild Type
KP Kras G12D Trp53 Δ/Δ
EKP Cdh1 KO Kras G12D Trp53 ΔIΔ
C
WT
KP
EKP
D
Diameter (µm) Passage 1
Diameter (µm) Passage 2
ns
WT (66)
KP (138)
EKP (124)
WT (65)
KP (127)
EKP (129)
E
F
MKI67 DAPI
G
CDH1 DAPI
H
P = 0.0013
P = 0.0284
P = 0.2233
% MKI67+ cells
50 µm
20 µm

Figure 5

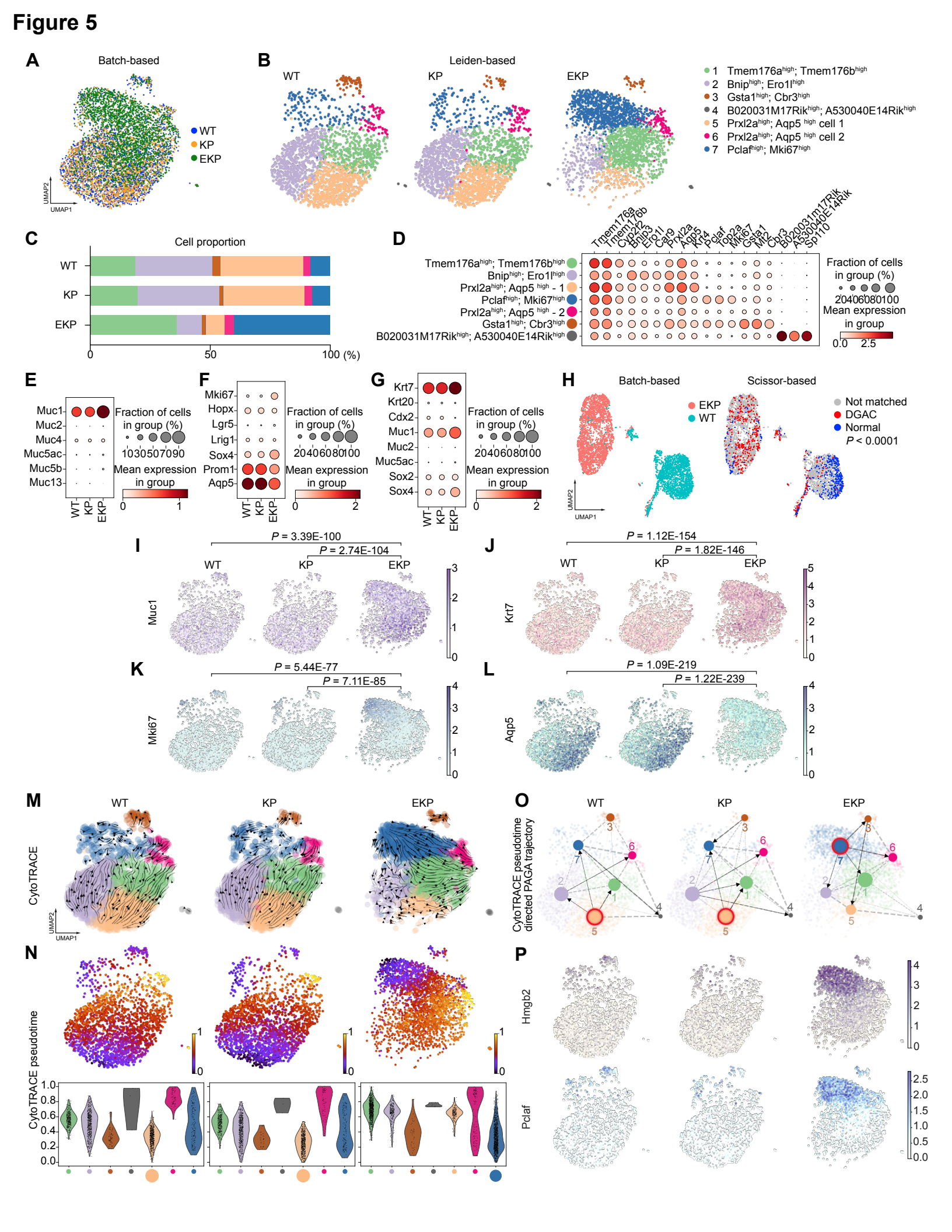

A
Batch-based
WT
KP
EKP
UMAP2
UMAP1
B
Leiden-based
WT
KP
EKP
1 Tmem176a^high; Tmem176b^high
2 Bnip^high; Ero1l^high
3 Gsta1^high; Cbr3^high
4 B020031M17Rik^high; A530040E14Rik^high
5 Prxl2a^high; Aqp5 ^high cell 1
6 Prxl2a^high; Aqp5 ^high cell 2
7 Pclaf^high; Mki67^high
C
Cell proportion
WT
KP
EKP
0
50
100 (%)
D
Tmem176a
Tmem176b
Cyp2f2
Bnip3
Ero1l
Car9
Prxl2a
Aqp5
Krt4
Pclaf
Top2a
Mki67
Gsta1
Mt2
Cbr3
B020031m17Rik
A530040E14Rik
Sp110
Tmem176a^high; Tmem176b^high
Bnip^high; Ero1l^high
Prxl2a^high; Aqp5 ^high - 1
Pclaf^high; Mki67^high
Prxl2a^high; Aqp5 ^high - 2
Gsta1^high; Cbr3^high
B020031M17Rik^high; A530040E14Rik^high
Fraction of cells in group (%)
20406080100
Mean expression in group
0.0
2.5
E
Muc1
Muc2
Muc4
Muc5ac
Muc5b
Muc13
Fraction of cells in group (%)
1030507090
Mean expression in group
0
1
WT
KP
EKP
F
Mki67
Hopx
Lgr5
Lrig1
Sox4
Prom1
Aqp5
Fraction of cells in group (%)
20406080100
Mean expression in group
0
2
WT
KP
EKP
G
Krt7
Krt20
Cdx2
Muc1
Muc2
Muc5ac
Sox2
Sox4
Fraction of cells in group (%)
20406080100
Mean expression in group
0
2
WT
KP
EKP
H
Batch-based
Scissor-based
EKP
WT
Not matched
DGAC
Normal
P < 0.0001
UMAP2
UMAP1
I
P = 3.39E-100
P = 2.74E-104
WT
KP
EKP
Muc1
3
2
1
0
J
P = 1.12E-154
P = 1.82E-146
WT
KP
EKP
Krt7
5
4
3
2
1
0
K
P = 5.44E-77
P = 7.11E-85
Mki67
4
3
2
1
0
L
P = 1.09E-219
P = 1.22E-239
Aqp5
4
3
2
1
0
M
WT
KP
EKP
CytoTRACE
UMAP2
UMAP1
O
WT
KP
EKP
CytoTRACE pseudotime directed PAGA trajectory
1
2
3
4
5
6
7
N
CytoTRACE pseudotime
1
0
CytoTRACE
1.0
0.8
0.6
0.4
0.2
0.0
P
Hmgb2
4
3
2
1
0
Pclaf
2.5
2.0
1.5
1.0
0.5
0.0

Figure 6

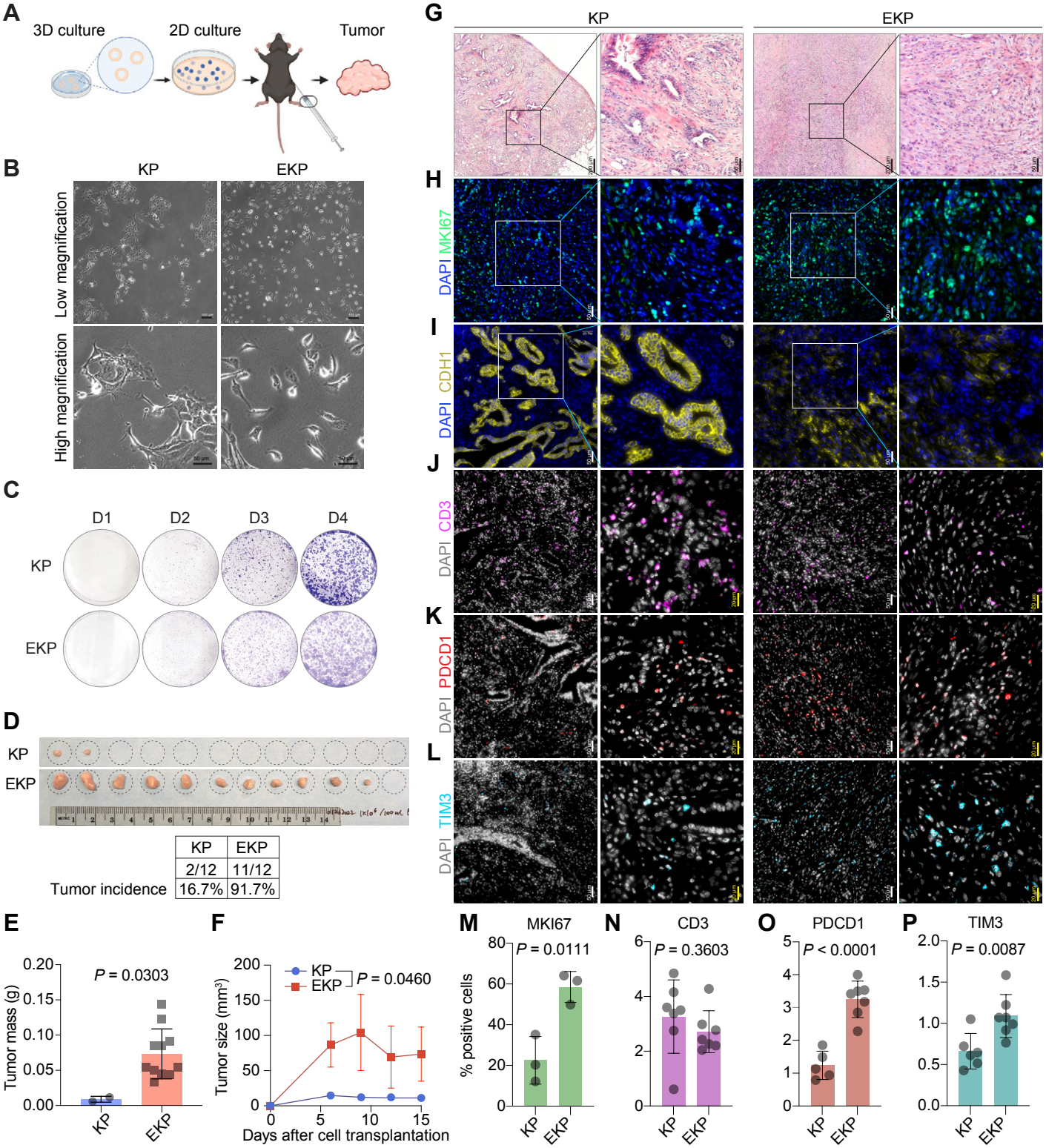
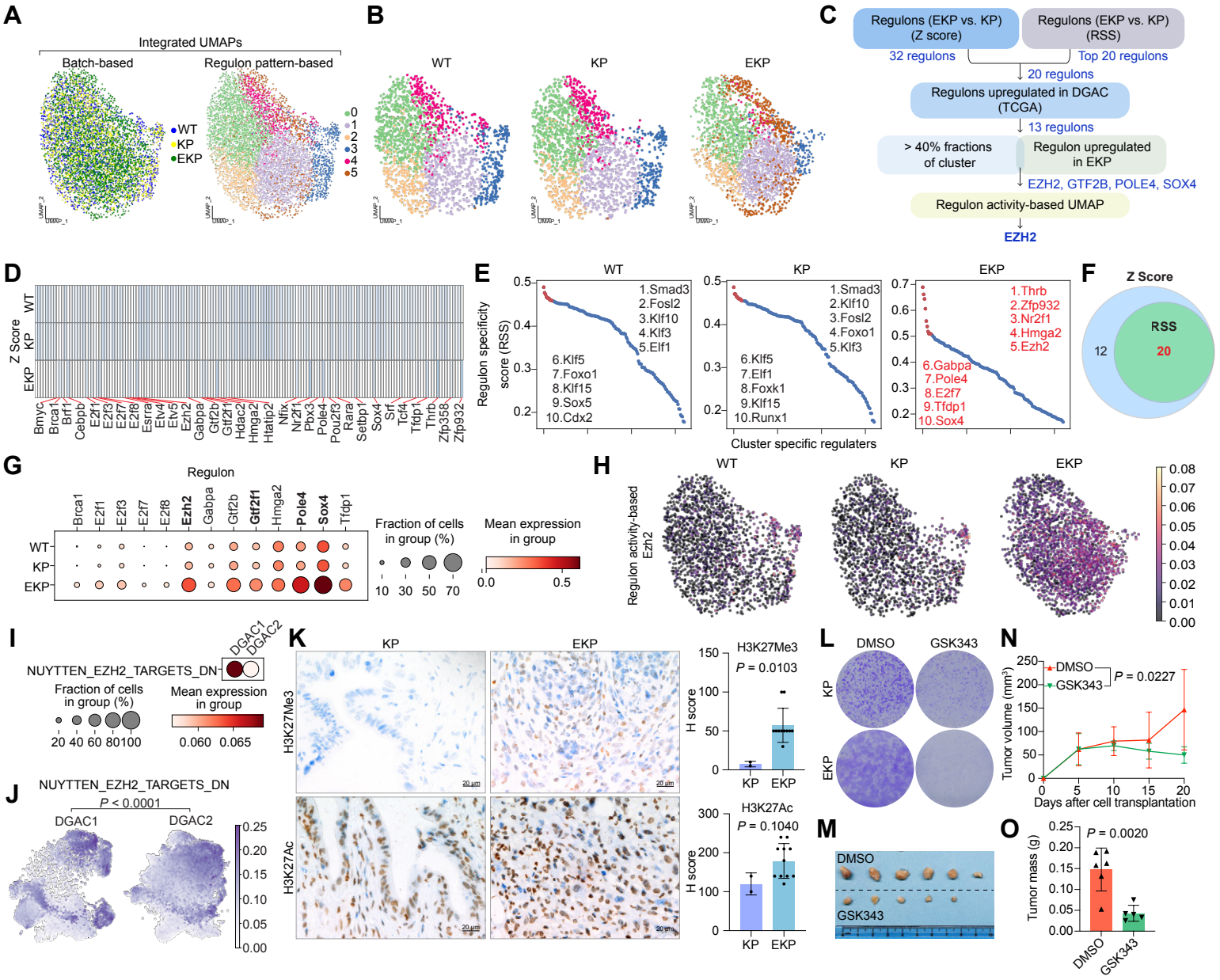
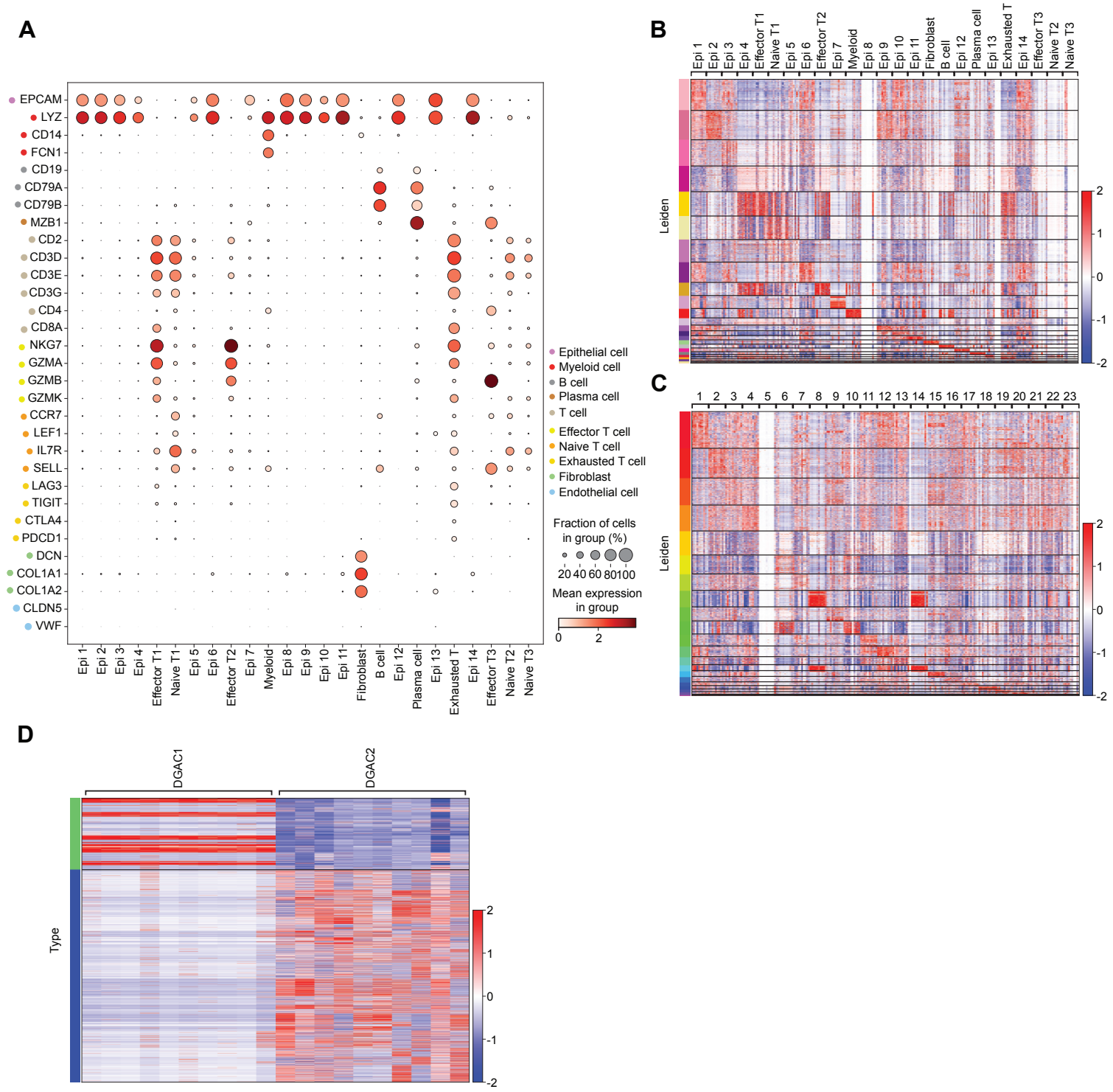

A
3D culture
2D culture
Tumor
B
KP
EKP
Low magnification
High magnification
C
D1
D2
D3
D4
KP
EKP
D
KP
EKP
Tumor incidence
KP
EKP
2/12
11/12
16.7%
91.7%
E
Tumor mass (g)
P = 0.0303
KP
EKP
F
Tumor size (mm³)
KP
EKP
P = 0.0460
Days after cell transplantation
G
KP
EKP
H
DAPI MKI67
I
DAPI CDH1
J
DAPI CD3
K
DAPI PDCD1
L
DAPI TIM3
M
MKI67
P = 0.0111
% positive cells
N
CD3
P = 0.3603
O
PDCD1
P < 0.0001
P
TIM3
P = 0.0087

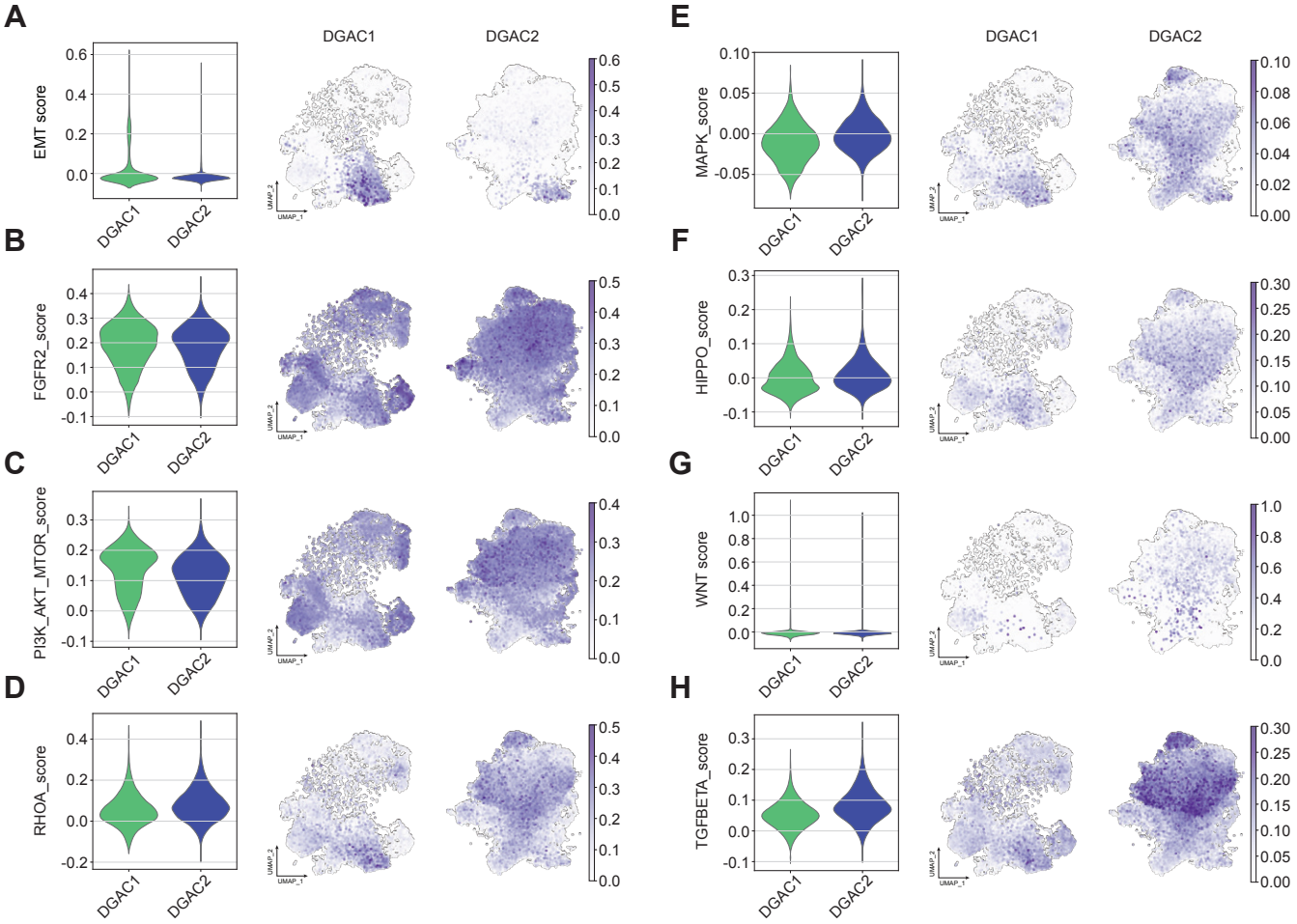
Figure 7


A
Integrated UMAPs
Batch-based
Regulon pattern-based
WT
KP
EKP
0
1
2
3
4
5
B
WT
KP
EKP
C
Regulons (EKP vs. KP) (Z score)
Regulons (EKP vs. KP) (RSS)
32 regulons
Top 20 regulons
20 regulons
Regulons upregulated in DGAC (TCGA)
13 regulons
> 40% fractions of cluster
Regulon upregulated in EKP
EZH2, GTF2B, POLE4, SOX4
Regulon activity-based UMAP
EZH2
D
Z Score
WT
KP
EKP
Bmyc
Brca1
Brf1
Cebpb
E2f1
E2f3
E2f7
E2f8
Esrra
Etv4
Etv5
Ezh2
Gabpa
Gtf2b
Gtf2f1
Hdac2
Hmga2
Htatip2
Nfix
Nr2f1
Pbx3
Pole4
Pou2f3
Rara
Setbp1
Sox4
Srf
Tcf4
Tfdp1
Thrb
Zfp358
Zfp932
E
WT
KP
EKP
Regulon specificity score (RSS)
1.Smad3
2.Fosl2
3.Klf10
4.Klf3
5.Elf1
6.Klf5
7.Foxo1
8.Klf15
9.Sox5
10.Cdx2
1.Smad3
2.Klf10
3.Fosl2
4.Foxo1
5.Klf3
6.Klf5
7.Elf1
8.Foxk1
9.Klf15
10.Runx1
1.Thrb
2.Zfp932
3.Nr2f1
4.Hmga2
5.Ezh2
6.Gabpa
7.Pole4
8.E2f7
9.Tfdp1
10.Sox4
Cluster specific regulaters
F
Z Score
RSS
12
20
G
Regulon
Brca1
E2f1
E2f3
E2f7
E2f8
Ezh2
Gabpa
Gtf2b
Gtf2f1
Hmga2
Pole4
Sox4
Tfdp1
WT
KP
EKP
Fraction of cells in group (%)
10
30
50
70
Mean expression in group
0.0
0.5
H
WT
KP
EKP
Regulon activity-based Ezh2
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0.00
I
DGAC1
DGAC2
NUYTTEN_EZH2_TARGETS_DN
Fraction of cells in group (%)
20
40
60
80
100
Mean expression in group
0.060
0.065
J
NUYTTEN_EZH2_TARGETS_DN
P < 0.0001
DGAC1
DGAC2
0.25
0.20
0.15
0.10
0.05
0.00
K
KP
EKP
H3K27Me3
H3K27Ac
H3K27Me3
P = 0.0103
H score
KP
EKP
H3K27Ac
P = 0.1040
H score
KP
EKP
L
DMSO
GSK343
KP
EKP
M
DMSO
GSK343
N
DMSO
GSK343
P = 0.0227
Tumor volume (mm³)
Days after cell transplantation
O
P = 0.0020
Tumor mass (g)
DMSO
GSK343

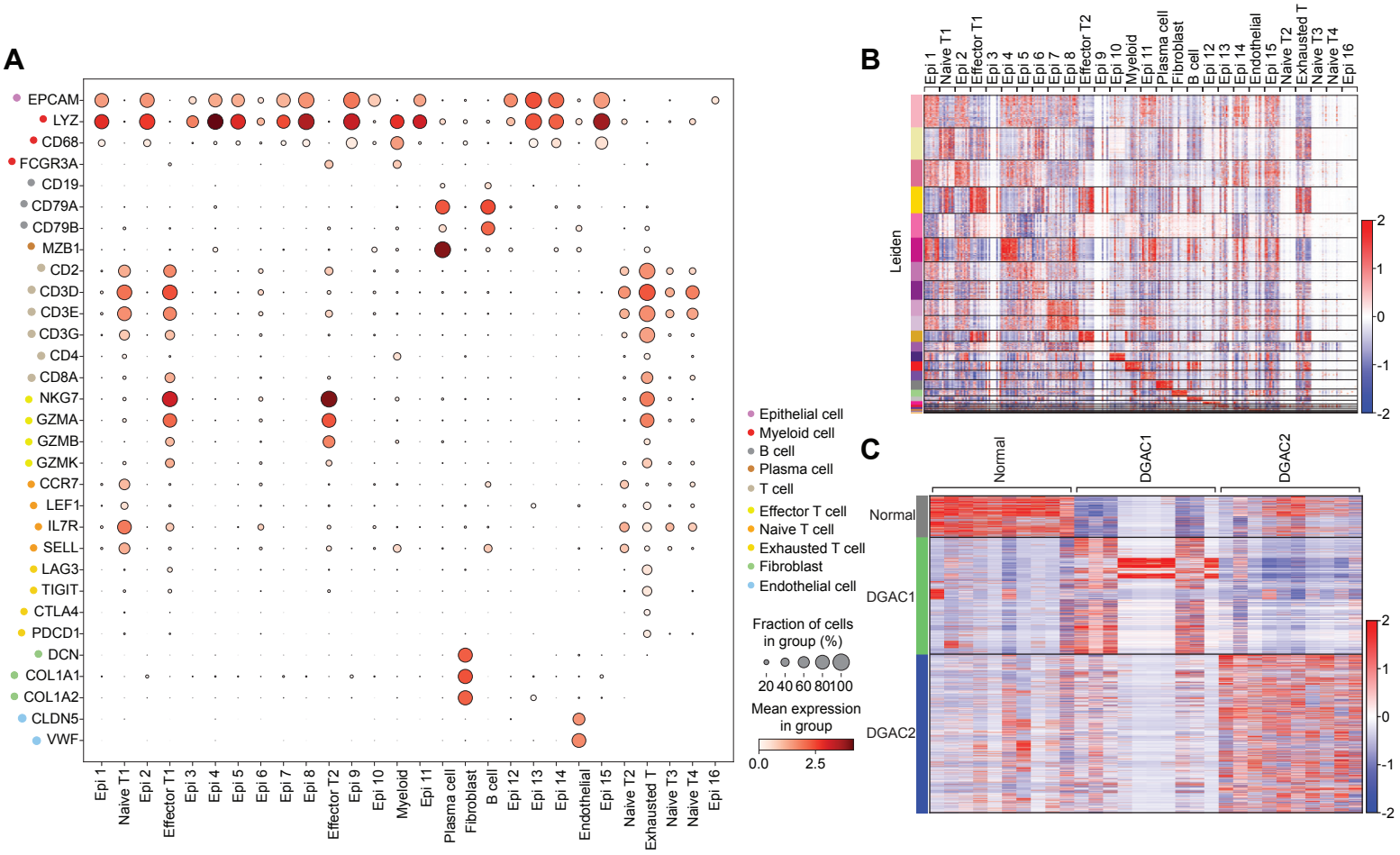
Supplementary Figure 1



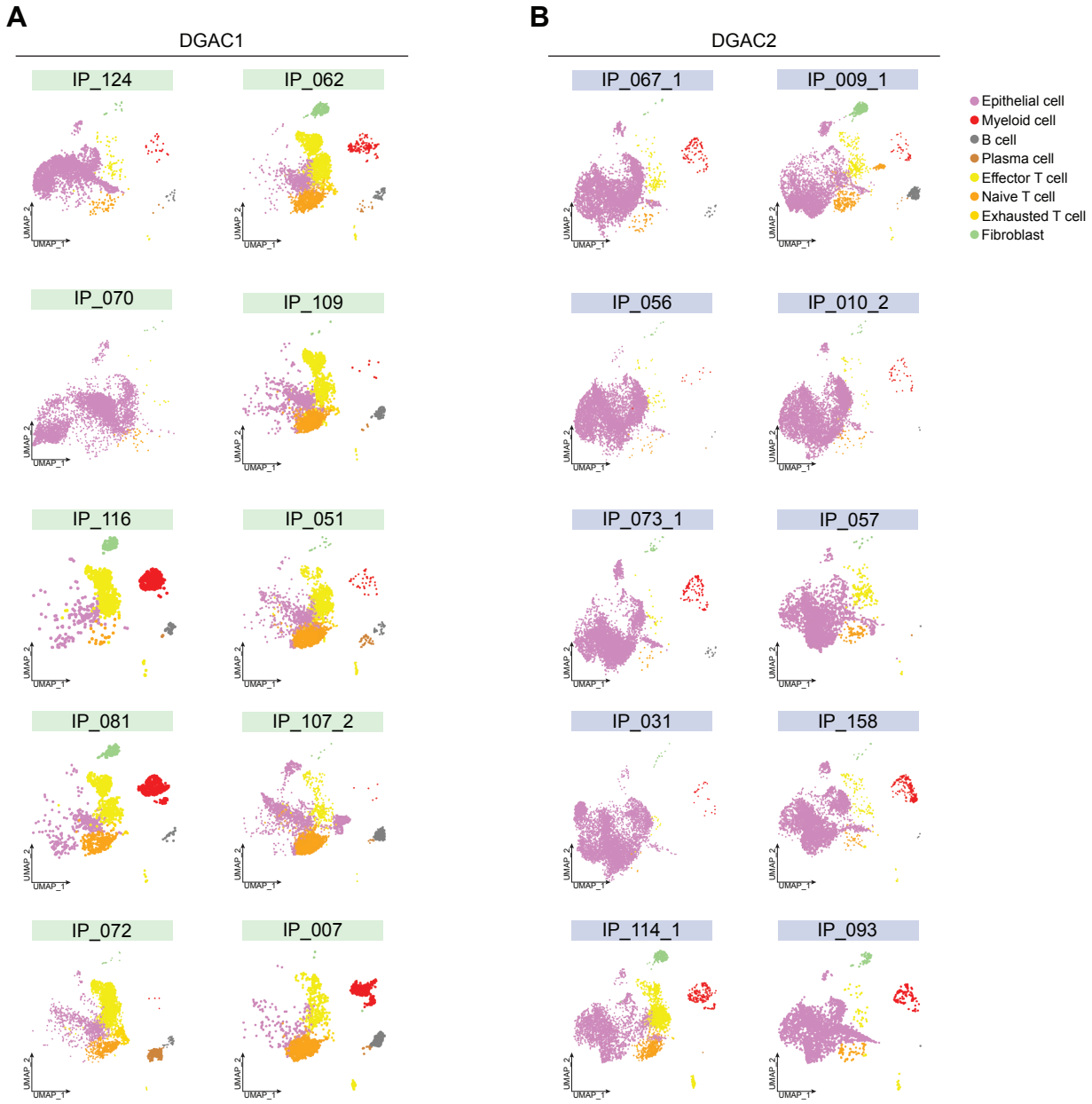
Supplementary Figure 2



Supplementary Figure 3



Supplementary Figure 4



Supplementary Figure 5

A

	Pathway	NES	pval	padj	
GOBP	GOBP_DIGESTION	2.08	5.1e-12	2.2e-09	DGAC2
	GOBP_CELL_JUNCTION_ASSEMBLY	1.71	3.5e-11	1.1e-08	
	GOBP_DIGESTIVE_SYSTEM_PROCESS	2.07	1.6e-10	3.8e-08	
	GOBP_CELL_CELL_JUNCTION_ORGANIZATION	1.88	1.9e-10	4.2e-08	
	GOBP_CELL_CELL_JUNCTION_ASSEMBLY	1.83	1.3e-07	1.3e-05	
	GOBP_EPIDERMIS_DEVELOPMENT	1.64	1.2e-07	1.3e-05	
	GOBP_OLEFINIC_COMPOUND_METABOLIC_PROCESS	1.82	1.4e-07	1.4e-05	
	GOBP_TIGHT_JUNCTION_ORGANIZATION	1.90	3.6e-07	3.3e-05	
	GOBP_RESPONSE_TO_WOUNDING	1.47	1.4e-06	1.1e-04	
	GOBP_PANCREAS_DEVELOPMENT	1.87	1.8e-06	1.4e-04	
	GOBP_POSITIVE_REGULATION_OF_LEUKOCYTE_CELL_CELL_ADHESION	**-2.06**	**6.0e-12**	**2.3e-09**	DGAC1
	GOBP_LEUKOCYTE_CELL_CELL_ADHESION	**-1.93**	**6.6e-12**	**2.3e-09**	
	GOBP_B_CELL_ACTIVATION	**-1.98**	**7.3e-13**	**3.5e-10**	
	GOBP_B_CELL_RECEPTOR_SIGNALING_PATHWAY	**-2.70**	**3.8e-15**	**2.1e-12**	
	GOBP_ADAPTIVE_IMMUNE_RESPONSE_BASED_ON_SOMATIC_RECOMBINATION_OF_IMMUNE_RECEPTORS_BUILT_FROM_IMMUNOGLOBULIN_SUPERFAMILY_DOMAINS	**-2.13**	**1.1e-15**	**7.2e-13**	
	GOBP_LYMPHOCYTE_MEDIATED_IMMUNITY	**-2.22**	**3.6e-16**	**3.3e-13**	
	GOBP_IMMUNE_RESPONSE_REGULATING_CELL_SURFACE_RECEPTOR_SIGNALING_PATHWAY	**-2.14**	**3.8e-16**	**3.3e-13**	
	GOBP_ACTIVATION_OF_IMMUNE_RESPONSE	**-2.20**	**4.2e-16**	**3.3e-13**	
	GOBP_ANTIGEN_RECEPTOR_MEDIATED_SIGNALING_PATHWAY	**-2.45**	**1.0e-16**	**2.0e-13**	
	GOBP_CYTOPLASMIC_TRANSLATION	**-3.25**	**8.2e-37**	**3.2e-33**	

B

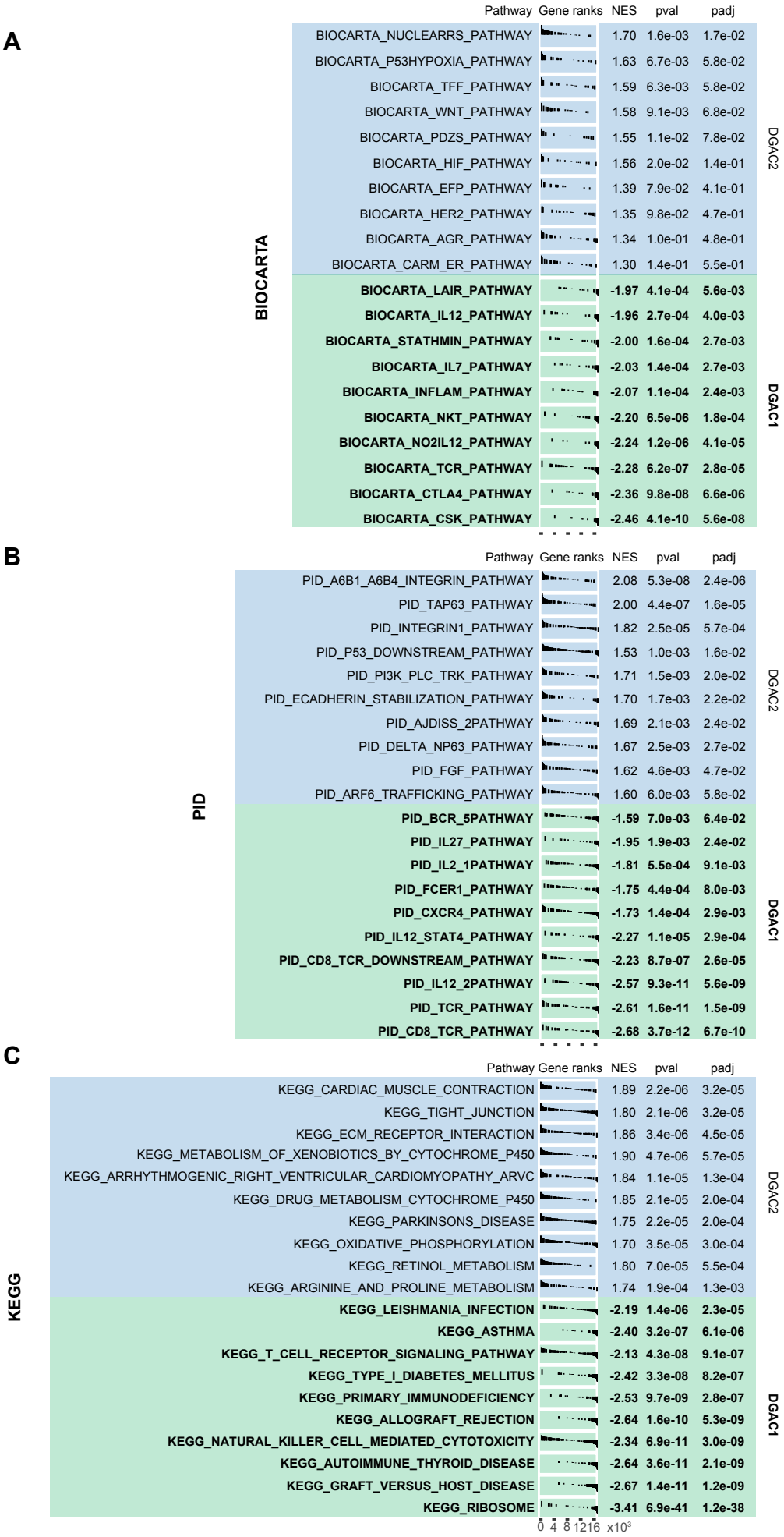
	Pathway	NES	pval	padj	
REACTOME	REACTOME_CELL_JUNCTION_ORGANIZATION	2.21	1.8e-13	1.1e-11	DGAC2
	REACTOME_CELL_CELL_COMMUNICATION	2.06	1.3e-11	6.9e-10	
	REACTOME_EXTRACELLULAR_MATRIX_ORGANIZATION	1.76	2.8e-10	1.4e-08	
	REACTOME_COLLAGEN_FORMATION	2.02	3.4e-09	1.6e-07	
	REACTOME_CELL_CELL_JUNCTION_ORGANIZATION	2.04	3.2e-08	1.4e-06	
	REACTOME_BIOLOGICAL_OXIDATIONS	1.78	6.0e-08	2.5e-06	
	REACTOME_PHASE_I_FUNCTIONALIZATION_OF_COMPOUNDS	1.92	9.4e-08	3.8e-06	
	REACTOME_FORMATION_OF_THE_CORNIFIED_ENVELOPE	1.93	1.2e-07	4.7e-06	
	REACTOME_DEGRADATION_OF_THE_EXTRACELLULAR_MATRIX	1.85	1.7e-07	6.2e-06	
	REACTOME_ASSEMBLY_OF_COLLAGEN_FIBRILS_AND_OTHER_MULTIMERIC_STRUCTURES	1.99	1.8e-07	6.5e-06	
	REACTOME_TRANSLATION	**-2.75**	**1.6e-31**	**1.5e-29**	DGAC1
	REACTOME_NONSENSE_MEDIATED_DECAY_NMD	**-3.27**	**3.4e-33**	**3.7e-31**	
	REACTOME_CELLULAR_RESPONSE_TO_STARVATION	**-3.19**	**5.1e-35**	**6.2e-33**	
	REACTOME_RESPONSE_OF_EIF2AK4_GCN2_TO_AMINO_ACID_DEFICIENCY	**-3.39**	**2.7e-36**	**3.7e-34**	
	REACTOME_SIGNALING_BY_ROBO_RECEPTORS	**-3.11**	**1.7e-36**	**2.8e-34**	
	REACTOME_REGULATION_OF_EXPRESSION_OF_SLITS_AND_ROBOS	**-3.15**	**4.9e-38**	**9.5e-36**	
	REACTOME_SELENOAMINO_ACID_METABOLISM	**-3.40**	**2.6e-38**	**6.4e-36**	
	REACTOME_EUKARYOTIC_TRANSLATION_INITIATION	**-3.36**	**6.8e-39**	**2.2e-36**	
	REACTOME_SRP_DEPENDENT_COTRANSLATIONAL_PROTEIN_TARGETING_TO_MEMBRANE	**-3.46**	**6.1e-43**	**3.0e-40**	
	REACTOME_EUKARYOTIC_TRANSLATION_ELONGATION	**-3.53**	**3.8e-45**	**3.7e-42**	

C

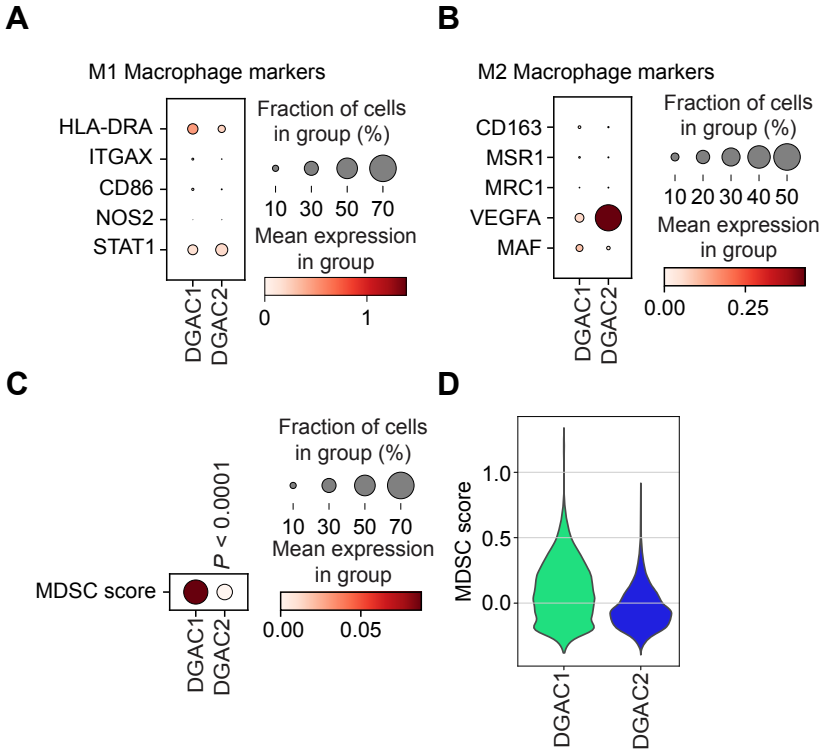
	Pathway	NES	pval	padj	
WP	WP_NUCLEAR_RECEPTORS_METAPATHWAY	1.70	2.1e-08	1.6e-06	DGAC2
	WP_NRF2_PATHWAY	1.85	2.6e-07	1.7e-05	
	WP_ARRHYTHMOGENIC_RIGHT_VENTRICULAR_CARDIOMYOPATHY	1.85	4.5e-06	1.9e-04	
	WP_FERROPTOSIS	1.77	1.0e-04	2.8e-03	
	WP_METAPATHWAY_BIOTRANSFORMATION_PHASE_I_AND_II	1.63	1.1e-04	2.9e-03	
	WP_EPITHELIAL_TO_MESENCHYMAL_TRANSITION_IN_COLORECTAL_CANCER	1.63	1.5e-04	3.4e-03	
	WP_HIPPOMERLIN_SIGNALING_DYSREGULATION	1.65	1.5e-04	3.4e-03	
	WP_OREXIN_RECEPTOR_PATHWAY	1.61	2.6e-04	5.5e-03	
	WP_ELECTRON_TRANSPORT_CHAIN_OXPHOS_SYSTEM_IN_MITOCHONDRIA	1.69	3.0e-04	5.7e-03	
	WP_TYPE_I_COLLAGEN_SYNTHESIS_IN_THE_CONTEXT_OF_OSTEOGENESIS_IMPERFECTA	1.78	4.1e-04	7.3e-03	
	WP_MICROGLIA_PATHOGEN_PHAGOCYTOSIS_PATHWAY	**-2.21**	**8.9e-06**	**3.1e-04**	DGAC1
	WP_CHEMOKINE_SIGNALING_PATHWAY	**-1.72**	**9.1e-06**	**3.1e-04**	
	WP_CANCER_IMMUNOTHERAPY_BY_PD1_BLOCKADE	**-2.33**	**3.0e-06**	**1.4e-04**	
	WP_TCELL_ANTIGEN_RECEPTOR_TCR_PATHWAY_DURING_STAPHYLOCOCCUS_AUREUS_INFECTION	**-2.11**	**1.6e-06**	**8.5e-05**	
	WP_TCELL_ACTIVATION_SARSCOV2	**-2.06**	**2.9e-07**	**1.7e-05**	
	WP_ALLOGRAFT_REJECTION	**-2.36**	**1.3e-09**	**1.3e-07**	
	WP_PATHOGENESIS_OF_SARSCOV2_MEDIATED_BY_NSP9NSP10_COMPLEX	**-2.58**	**9.6e-11**	**1.1e-08**	
	WP_TCELL_RECEPTOR_TCR_SIGNALING_PATHWAY	**-2.46**	**1.6e-11**	**2.5e-09**	
	WP_MODULATORS_OF_TCR_SIGNALING_AND_T_CELL_ACTIVATION	**-2.58**	**3.5e-12**	**8.3e-10**	
	WP_CYTOPLASMIC_RIBOSOMAL_PROTEINS	**-3.46**	**1.9e-45**	**8.9e-43**	

Gene ranks: 0 4 8 12 16 x10^{3}

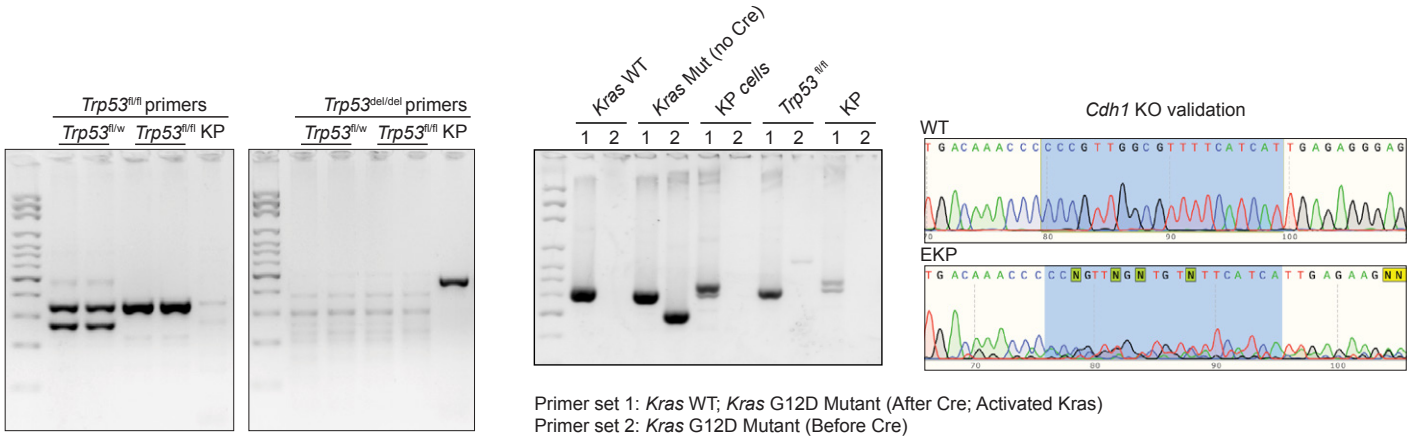
Supplementary Figure 6



Supplementary Figure 7

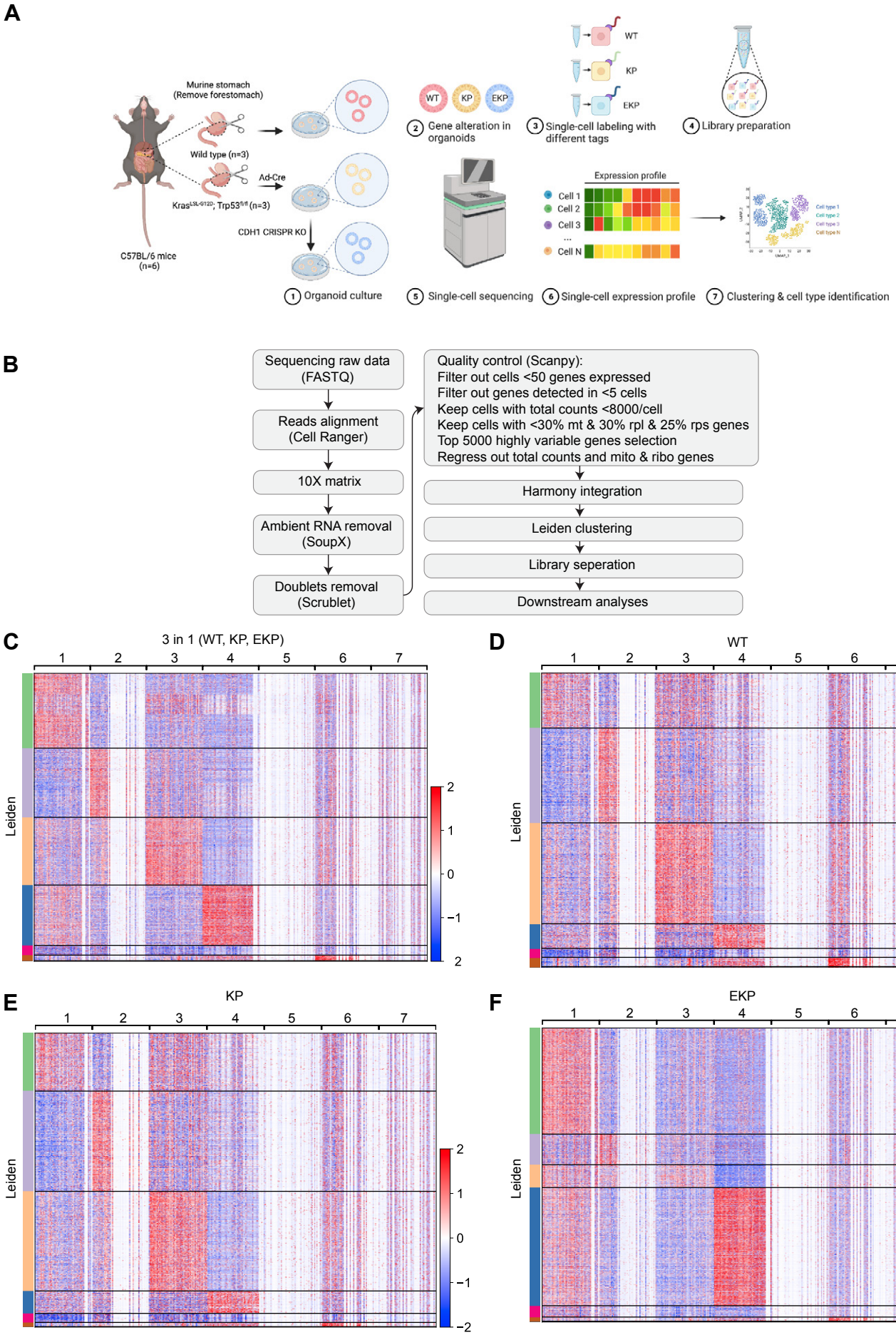


Supplementary Figure 8



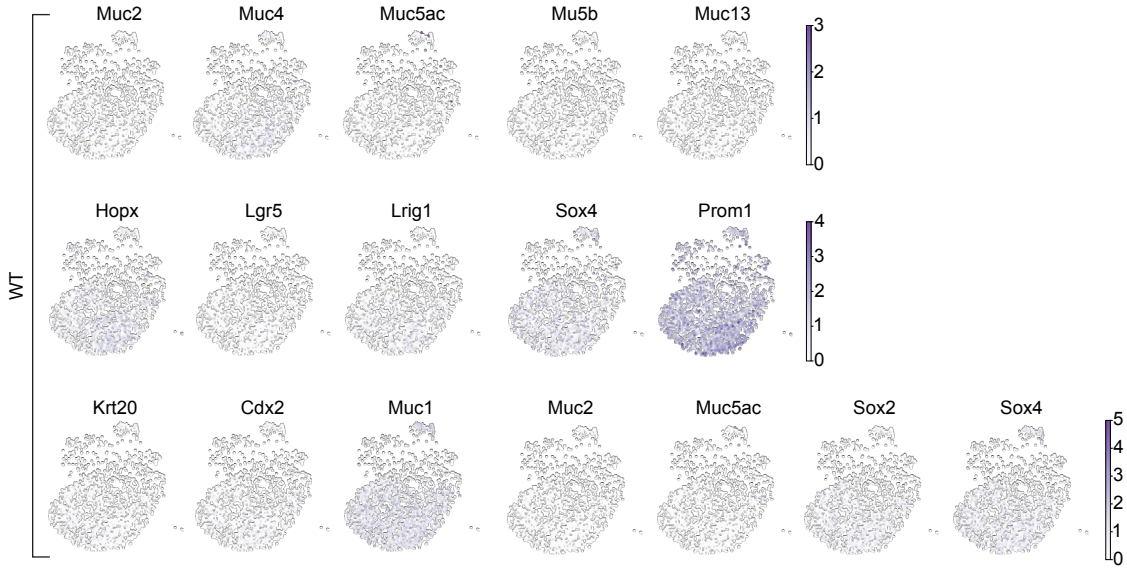
Primer set 1: *Kras* WT; *Kras* G12D Mutant (After Cre; Activated Kras)
Primer set 2: *Kras* G12D Mutant (Before Cre)

Supplementary Figure 9

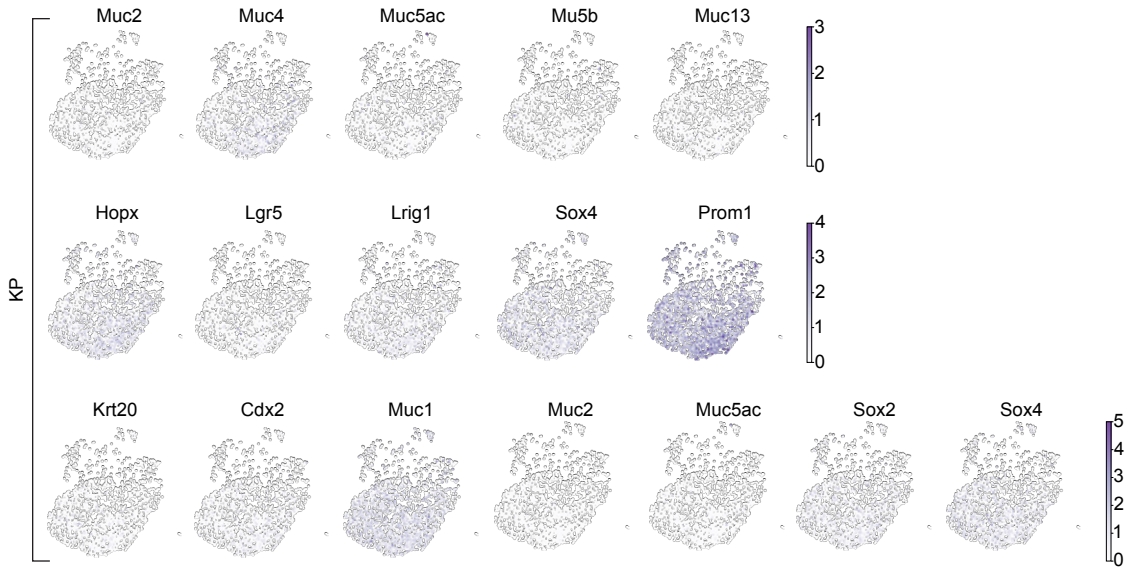


Supplementary Figure 10

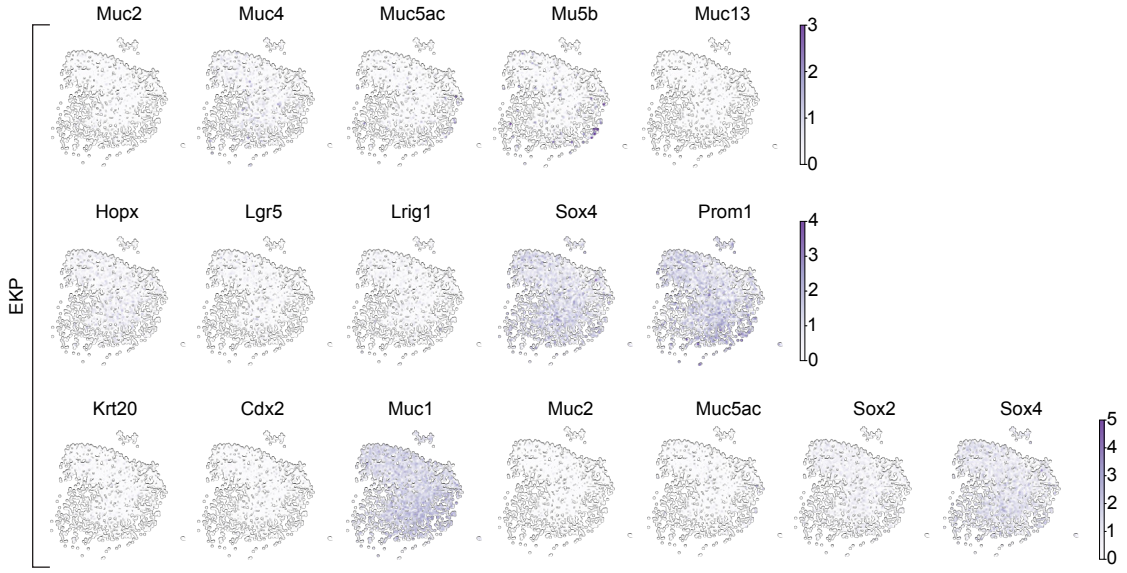
A



B



C

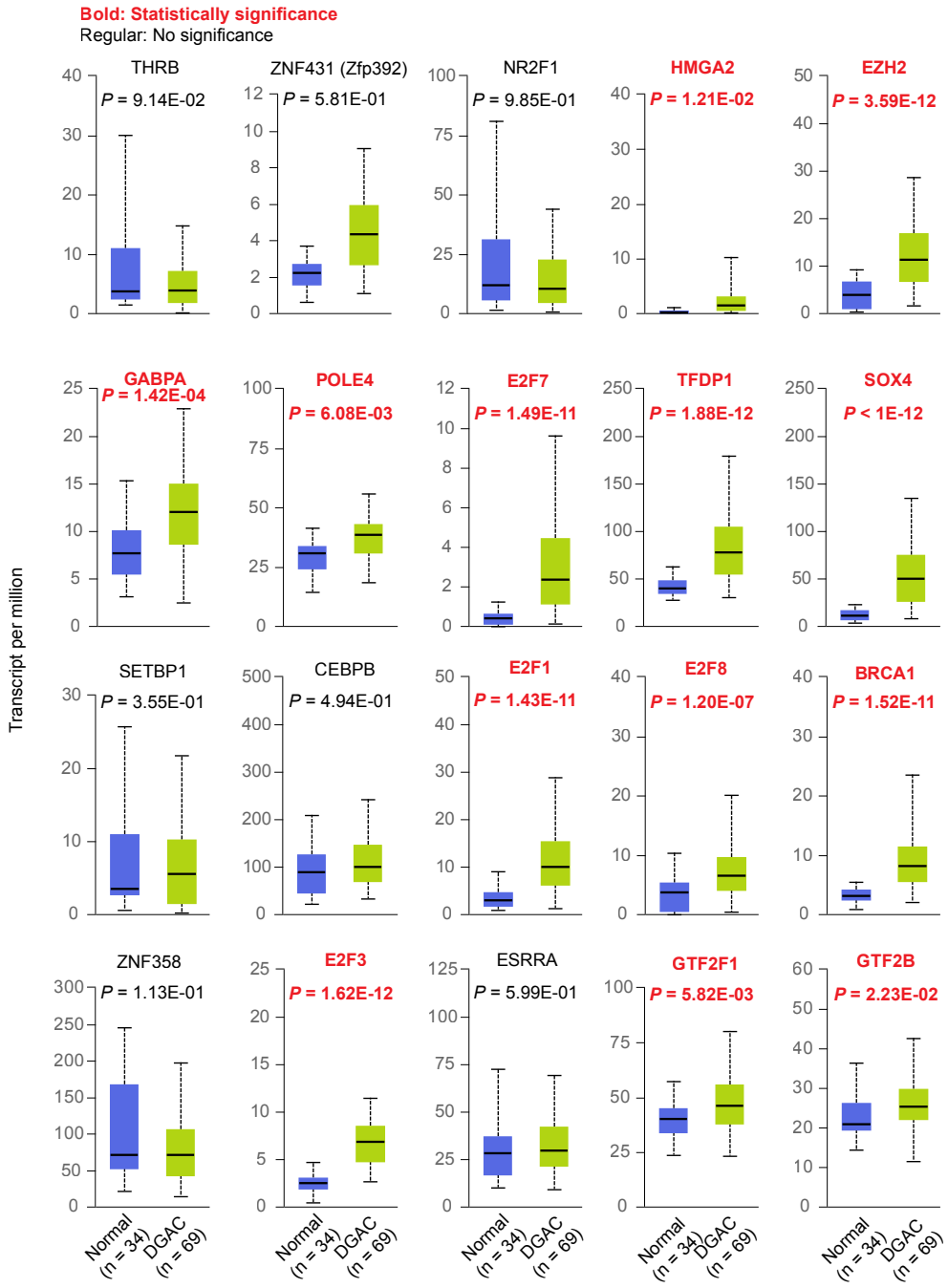
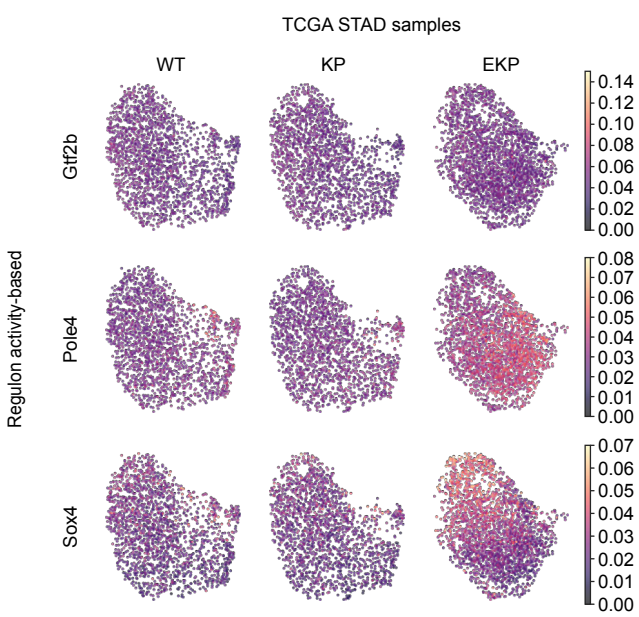


D

P value	EKP vs WT	EKP vs KP
Muc2	0.8679	0.8165
Muc4	0.9837	0.9994
Muc5ac	0.2626	0.2042
Muc5b	0.0080	0.0105
Muc13	0.9813	0.9036
Hopx	2.82E-16	8.96E-14
Lgr5	0.0343	0.0336
Lrig1	0.0002	0.0001
Sox4	6.21E-60	2.36E-63
Prom1	9.52E-131	3.49E-131
Krt20	0.3123	0.3643
Cdx2	4.42E-13	2.81E-13
Sox2	0.1350	0.5387

Supplementary Figure 11

A



B