Detailed Survey of an in-vitro Intestinal Epithelium Model by Single-Cell Transcriptomics

Ran Ran[1, 3]; Javier Munoz[1,2], Smrutiti Jena[1], Leopold N. Green[1], Douglas K. Brubaker*[1, 3]
[1] Weldon School of Biomedical Engineering, Purdue University; [2] Purdue Interdisciplinary Life Science Program. [3] Current Affiliation: Department of Pathology, Case Western Reserve University
*Correspondence: dkb50@case.edu.

Abstract

The gut plays a critical role in maintaining human health by facilitating the absorption of nutrients, regulating metabolism, and interacting with the immune system and gut microbiota. The co-culture of two human colorectal cancer cell lines, Caco-2 and HT29, on Transwell is commonly used as an *in vitro* gut mimic in studies of intestinal absorption pharmacokinetics, gut mechanics, and gut-microbe interplay given the similar morphology, expression of transporters and enzymes, and barrier function. However, to sufficiently evaluate the translatability of insights from such a system to human physiological contexts, a detailed survey of cell type heterogeneity in the system and a holistic comparison with human physiology are needed to be conducted rather than by the presence of a few well-studied proteins. Single-cell RNA sequencing provides high-resolution expression profiles of cells in the co-culture, enabling the heterogeneity to be characterized and the similarity to human epithelial cells to be evaluated. Transcriptional profiles of 16019 genes in 13784 cells were acquired and compared to human epithelial cells (GSE185224). We identified the intestinal stem cell-, transit amplifying-, enterocyte-, goblet cell-, and enteroendocrine-like cells together with differentiating HT29 cells in the system based on the expression of canonical markers in healthy adult human epithelial cells. The epithelium-like co-culture was fetal intestine-like, with less variety of gene expression compared to the human gut. Transporters for major types of substance (lipid, amino acid, ion, water, etc.) were found transcribed in the majority of the enterocytes-like cells in the system. However, some of the well-studied transporters such as FATP4 and GLUT2 were absent. Toll-like receptors were not highly expressed in the sample, yet the treatment of lipopolysaccharide still caused a mild change in trans-epithelial electrical resistance and gene expression, possibly by the interaction with CD14, the co-receptor for TLRs. Overall, the Caco-2/HT29 co-culture is a cost-effective epithelium model for drug permeability testing or mechanical simulation, but its phenotypic discrepancy with the real epithelium is not negligible. As a result, its response to biological factors might not provide transferrable knowledge to the study of human gut physiology, especially the innate immune aspect.

Introduction

Inflammatory Bowel Diseases (IBD) are complex and debilitating diseases that affect millions of people worldwide[1]. These disorders are characterized by an exaggerated immune response that leads to chronic inflammation in the gastrointestinal tract, resulting in a range of symptoms including abdominal pain, diarrhea, and rectal bleeding[2]. The two most common forms of IBD are Crohn's disease (CD), which can affect any part of the digestive tract and is characterized by transmural inflammation, and Ulcerative Colitis (UC), which is primarily confined to the colon and rectum and is characterized by continuous inflammation of the inner lining of the colon[2]. The prevalence of IBD and colorectal cancer in the United States has been increasing dramatically in recent years[3]. According to recent estimates, approximately 3 million adults in the US have been diagnosed with IBD, with an estimated annual healthcare cost of over $6 billion[3]. This scenario is exacerbated by the higher risk of IBD patients relapsing within 1 year (28.7%) and 2 years (38.4%) after therapeutic de-

escalation[4]. Additionally, studies have shown that individuals with IBD have an increased risk of developing colorectal cancer, with the risk increasing with the duration and extent of the disease[5]. Clearly, IBD is a significant health problem that requires urgent attention. Understanding the biology of IBD and the interactions between the immune system, the epithelium, and the microbiome is critical for the development of effective therapies for this disease[2].

Caco-2 cells, derived from human colorectal adenocarcinoma, have been extensively employed as a remarkable proxy for the intricate gut epithelium[6]. The salient features that render these cells an exemplary in vitro model include their ability to differentiate into enterocyte-like cells, recapitulating the intestinal absorptive characteristics[6]. Furthermore, Caco-2 cells exhibit brush border enzymes, tight junctions, and a plethora of nutrient transporters, providing an accurate reflection of the physiological environment[6]. Additionally, their robust and reproducible barrier function makes them indispensable for scrutinizing drug permeability and absorption studies[7]. Lastly, the amenability of Caco-2 cells to various experimental manipulations bolsters their utility in exploring the multifaceted interactions between the gut epithelium and dietary components, pathogens, or pharmaceutical agents[6]. On the other hand, HT29 cells, another human colon carcinoma cell line, have garnered significant attention as a valuable model for mimicking gut epithelial cells[8]. These versatile cells exhibit the ability to differentiate into diverse intestinal cell types, including goblet and enteroendocrine cells, thus providing an authentic representation of the intestinal cellular milieu[8]. Moreover, HT29 cells contribute to the understanding of mucus secretion dynamics, which play a critical role in maintaining gut barrier integrity and host defense[9].

However, to sufficiently evaluate the translatability of insights from such a system to human physiological contexts, a detailed survey of cell type heterogeneity in the system and a holistic comparison with human physiology need to be conducted, not just by the presence of a few well-studied proteins[10]. Single-cell RNA sequencing (scRNA-seq) represents a cutting-edge approach to meticulously assess the resemblance between in vitro gut mimics, such as Caco-2 and HT29 cells, and the authentic human intestinal epithelium[11]. By offering a high-resolution transcriptomic landscape of individual cells, scRNA-seq facilitates the identification of distinct cell populations, thus enabling a comprehensive comparison of cellular composition and heterogeneity between these models and native tissue[11]. Furthermore, this powerful technique uncovers the dynamics of gene expression and operative signaling pathways in both systems, shedding light on their functional congruence[11]. Employing scRNA-seq in this context also aids in the discovery of potential limitations or disparities within in vitro models, fostering the refinement and optimization of these systems for enhanced physiological relevance[12]. As a result, the subsequent improvements in these models pave the way for more accurate and insightful investigations of gut epithelial interactions with external factors, ultimately advancing our understanding of human gastrointestinal physiology and disease[10]. In this study, we employed scRNA-seq to investigate the molecular signatures of the co-culture and assess its potential for producing results that can be generalized to the human gut.

Result

Determination of Human Cell Type-like Signatures Represented in the Germ-Free Co-culture Model

We cultured Caco-2 (ATCC, passage no. 8) and HT29 (ATCC, passage no.13) separately at 1x105 cells/ cm2 and seeded them together into the apical chamber of three Transwells (Corning Inc, Costar, USA) with a ratio of 9:1 on day 0. Two Transwells were used as

controls, where cells were co-cultured and harvested on day 23. The other Transwell was treated with 1 ng/ml Escherichia coli O111:B4 lipopolysaccharides on day 22 and harvested on day 23. Then all samples were prepared for single-cell RNA sequencing as per the manufacturer's instructions (10X genomics) (Fig. 1A). Cell morphology agreed with the previous description of the Caco-2/HT29 monolayers (Fig. S1A-B). The TEER values reported from control treatments were within the range of values reported by Hoffmann et al.[13], indicating that the established Caco-2/HT29 co-culture system possesses properties comparable to those in other reported cases (Fig. S1C).

Cell identities were annotated based on their expression of previously reported signatures (ISC: LGR5, ASCL2, OLFM4, SMOC2, OLFM4[14]; TA: MKI67, TOP2A[15]; EC: APOA4, ANPEP, FABP2[16]; GC: MUC2[17]; EEC: PAX4, INSM1, CHGA, CHGB[18]) in healthy human intestinal epithelial cells (Fig. 1B, E). The general cell types were subdivided by their stemness/proliferating state (Fig. 1C) and were further segmented into more detailed cell groups (Fig. 1D) based on uniquely expressed genes (Fig. 1F). Cell subsets with no similar known human analogy were annotated based on their cancer-related gene expression and differentiation state. Overall, the Caco-2/HT29 co-culture system was highly heterogeneous. Quantitatively, about 6% of the co-cultured cells were intestinal stem cell-like cells (ISCLC), 30% were transit amplifying-like cells (TALC), 45% were enterocyte-like cells (ECLC), 1% were secretory progenitor-like cells, 2% were early enteroendocrine cell-like cells (EECLC), 1% were goblet cell-like cells (GCLC), and 15% were undifferentiated/differentiating HT29 cells (U-/D-HT29) (Fig. 1G). There was also a small cluster that exhibited fetal-like features, like the expression of TFAP2C (Table S1). All clusters had comparable sequencing depth (Fig. S2A) and mitochondrial gene percentages (Fig. S2B). The differentially expressed genes in each cluster were reported in Table S1. The conjectured origins of cell types, namely the monoculture Caco-2 or HT29, were determined by the overlap between cell type signatures and Caco-2/HT29 signatures (Table S2), identified from the bulk RNA-seq data of these two cell types provided by the Cancer Cell Line Encyclopedia (CCLE). Trajectory analysis of the system showed two distinct developmental branches, absorptive-like lineage ISCLC-TALC-ECLC and HT29/secretory-like lineages HT29-SECPLC-GC/Early ECLC (Fig. 1I). The arrangement of cluster nodes along the PAGA map followed the manually annotated time order.

We then asked how well the Caco-2/HT29 co-culture system can recapitulate the physiology of the human gut epithelium. We examined the expression profile of each major cell type in the model. Acknowledging that Caco-2 and HT29 were derived from colorectal adenocarcinoma, we also interrogated the cancer phenotypes of cells and asked if any of these properties disqualified the co-culture model for mimicking human gut epithelium.

Characteristics of the stem and proliferative populations

Healthy human gut epithelium self-renews every 4-5 days. This quick turnover is initiated by the differentiation of ISCs located at the base of the crypt. Specifically, ISCs move upwards along the crypt-villi axis, leaving their stem niche and entering the region called transit-amplifying zone, where they rapidly proliferate and commit to either absorptive or secretory fate while transiting to the villus tip[19,20]. In the co-culture, we identified 2 ISCLC subgroups, 2 early TALC subgroups, 1 intermediate TALC group, and 4 TALC subgroups (Fig. 2A). Noticeably, ISC and TA marker ASCL2, were universally expressed in all undifferentiated cells, which also agreed with the highly proliferative phenotype of cancer cells (Fig 2E-F, Fig. 2B).

Both ISCLCs had a fraction of cells expressing classic intestinal stem markers LGR5 and SMOC2[14] (Fig. 2B). A few cells from the ISCLCs were also expressing cancer stem cells (CSCs) marker PROM1 (CD133)[21], fetal/cancer marker alpha-fetoprotein (AFP)[22,23] and its homologous albumin (ALB), and proliferative marker MKI67 and TOP2A at a relatively low level (Fig. 2C). ISCLC 1 uniquely expressed PLA2G2A, a phospholipase A2 family member that is highly expressed in fetal ISC[24] and adult human Paneth cells (PC)[25]. PLA2G2A is known to be a Wnt inhibitor and PC differentiation modulator[26], so its expression might indicate the Wnt activities and/or the identity of fetal small intestine ISCs or adult PCs. However, ISCLC 1 did not express PC markers LYZ[27], DEFA4/5[27], or PRSS2[28] (Fig. S2C), so it was unlikely to be adult PC-like cells or the stem-like that have been reported to appear after inflammation in a previous study[29]. Given the ISCLC 1 also expressed ISC marker AXIN2[30], CSC marker and embryonic development regulator CD24[31], SOX2[32], and BMP4[33], it was most likely to be fetal ISCs with Wnt inhibition activities. By comparison, ISCLC 2 featured the strong expression of fetal ISC marker FGG while having little PLA2G2A. Since Wnt signaling maintains the self-renewal of ISCs[34], the active transcription of the Wnt inhibitor PLA2G2A in ISCLC1 might indicate it was at the beginning of differentiation and cue the loss of stemness in this population while ISCLC2 was at a more undifferentiated state. Overall, the expression profiles of both ISCLC subgroups resembled those observed in fetal cells and cancer cells.

Two early TALCs had decreased ISC marker expressions and increased transcription of genes associated with proliferation such as MKI67 and TOP2B. Compared to early TALC-1, early TALC-2 still had a small fraction of cells that remained ISC/CSC markers positive. Meanwhile, this subgroup also highly expressed DKK1, a putative Wnt inhibitor[35], and YAP's downstream ANKRD1[36]. The potential Wnt inhibiting effect of DKK1 might signal the loss of stemness, thus agreed with the assigned TALC identity. In fact, the expression of DKK1 was negatively correlated with the stem regulator AGR2 and positively correlated with solute carrier family 2 member 3 SLC2A3 (GLUT3), indicating the decreased stemness and increased metabolic activities.

The activities of Wnt inhibitors in the stem/proliferative populations invite further investigation in the role of Wnt signaling pathway in the co-culture maintenance. We found the expression of genes that were engaged in the Wnt signaling, such as frizzled class receptor FZD 2/3/4/5/6/7[37] and transcription factor TCF4[38], as well as the transcription of several Wnt downstream genes like ASCL2, MYC[39], ID1[40], HES1[41], SOX9[42], etc., which means the co-cultured cells were universally eligible of receiving Wnt signals and proceeding the downstream transcription. Despite the broad expression of Wnt-associated genes, Wnt ligands themselves were not highly expressed. Only Wnt11 was transcribed in a faction of the co-cultured cells and Wnt10A in a few U-HT29 and D-HT29 cells at a low level (Fig. S3A). Such observation is reasonable, considering Paneth and mesenchymal cells are the main producers of Wnts in healthy human gut[43], and the co-culture shares similarity with gut epithelium. Still, it raised the question how the Wnt pathway could be activated with such a limited Wnt secretion without the addition of exogenous Wnt (no Wnt protein in the culture medium or LPS treatment). It is possible that the secretion of Wnt11 activates non-canonical Wnt signaling[44], the mutation on the downstream components such as APC/beta-catenin (which is well-characterized in both Caco-2 and HT29 cell lines[45,46]), and/or the crosstalk between Wnt and other signaling pathways, such as TGF beta signaling[47].

The ISC/CSC markers were further diminished in Intermediate TALC and TALC (Fig. 2B). TALC-1 featured high proliferative marker expression. The expression of AFP and ALB in

TALC-2 was indicative of the resemblance of its features to the TA cells found in the fetal small intestine[22,23]. This finding further supports the notion that the epithelium-like tissue derived from Caco-2/HT29 exhibits characteristics akin to the fetal intestine. Nevertheless, such finding was anticipated given the reported embryonic genes re-expressed or upregulated in colorectal adenocarcinoma[48]. TALC-4 also exhibited high levels of MKI67 and TOP2A, but uniquely expressed gonadotropin-releasing hormone 2 (GNRH2), a hormone with unclear function. Unlike GNRH1, GNRH2 is commonly expressed in a variety of human peripheral tissues but is seldom found in brain regions associated with gonadotropin secretion[49], suggesting it might not function as a hormone in the same way GNRH1 does. Furthermore, it was observed that GNRH2 had an anti-proliferative effect on prostate, ovarian, breast, and endometrial cancer cells[50]. TALC-3 uniquely expressed Zinc Finger Protein 90 (ZNF90), a gene whose function is unknown but predicted to enable DNA-binding transcription factor activity[51].

We examined the expression of previously reported human ISC/TA signatures in the model[14]. Although LGR5 and SMOC2 were expressed, there was no active transcription of human adult ISC marker genes OLFM4, RARRES2, or CD44 (Fig. 2B, Fig. 2D). Interestingly, the lack of OLFM4 is a feature of the immature fetal intestine, strengthening the fetal-like features of the co-culture[52]. However, we did not find any BEX5+ uniform progenitors[24] that are reported to occur during the first trimester (Fig. S2C), suggesting the features were not early fetal-like. Overall, some of the human ISC differentially expressed genes (DEGs) were not expressed in the ISCLC, and conversely, only a few of the ISCLC DEGs were expressed in human ISC (Fig. S4A). A total of 1177 genes were expressed in over 30% of the population in both ISCLC and human ISC, whereas 134 genes expressed in ISCLC were not commonly expressed in human ISC. Additionally, 2258 genes commonly expressed in human ISC were not commonly expressed in ISCLC (Fig. 2D). By comparison, TALCs in the co-culture model shared more similarity with the human TA cells. Proliferative genes such as mini chromosome maintenance genes (MCMs) and cell division cycle (CDCs), which were differentially expressed in human TAs, were also prevalent in TALCs (Fig. 2F). All DEGs and 1443 out of 1566 commonly expressed genes found in TALCs were also expressed in human TAs (Fig. 2G, S4B).

Enterocyte-like cells and their absorption preference

Enterocytes express a wide array of transporters, which play crucial roles in the absorption and secretion of nutrients, ions, and drugs, thereby maintaining the integrity of the intestinal barrier[53]. These transporters can be broadly classified into families such as ATP-binding cassette transporters, solute carrier transporters, and ion channels, each with distinct functions and substrate specificities. The activity of these transporters is tightly regulated and can be affected by various factors, including genetic variations, diet, and medication use. Caco-2/HT29 system is a popular high-throughput model to study the drug intestinal absorption. This system performs well in testing drugs that are absorbed passively, and a variety of transporters have also been identified on the surface of the Caco-2 cells[54]. The comprehensive profiling of transporters on Caco-2/HT29 remains incomplete due to the limited characterization methods available for the co-culture system during its establishment. Here, we identified 7 early ECLC populations and 1 mature ECLC population that resembles small intestine enterocytes (Fig. 3A). Their ECLC identity was defined by the expression of major mature enterocyte transporters APOA4, ANPEP, and FABP2[53,55,56]. The degree of their differentiation was annotated based on the expression of proliferative markers in these cells (Fig. 3B). We queried the expression of 9 categories of functional proteins that are present and responsible for digestion and absorption in healthy human enterocytes (Fig. 3C). The

signature panel revealed unique features for each cell type, thus demonstrating the heterogeneity within the ECLC population. Adjacent to the fetal-like TALC-2 on UMAP, early ECLC-2 was actively transcribing AFP and ALB. Similarly, ECLC-4 featured a high expression of GNRH2. Early ECLC-6 exhibited undifferentiated HT29 features (i.e., LYZ, CD24, CD44, BMP4) at a lower level compared to the identified U-HT29 group (Fig. S2C) and was also positive for several SI enterocytes signature genes (Fig. 3B-C).

Overall, all ECLCs expressed a reduced spectrum of transporters and enzymes compared to the human enterocytes. Observed in co-cultured cells were fatty acid transporters (APOA1/4, APOC3), digestive enzyme (ANPEP, DPP4), vitamin retinol/retinaldehyde metabolism RBP2/4, and water channel AQP3/10/11 (Fig. 3C). "They also actively transcribed several transporters including the drug transporter ABCC2, the bile acid transporter ABCC3, and the cation transporter SLCO2B1 (OATP-B). Notably, they also transcribed SLC2A3 (GLUT3), a glucose transporter whose activities are usually restricted to the brain but has also been detected in the placenta[57] and CRC tissue[58] (Fig. 3C). The GLUT3 activity increases the glycolytic capacity of cancer cells to fulfill their needs for rapid growth and metastasis[58,59]. In the co-culture system, ECLC-3 and 5 expressed significantly higher GLUT3 than others.

However, the co-culture system lacked the expression of some well-studied essential transporters, such as SLC27A4[60] (FATP4) or CD36[61] for long-chain fatty acid absorption, ABCG5/G8 or NPC1L1 for cholesterol absorption[62], SLC5A1 (SGLT1) or SLC2A2 (GLUT2) for glucose absorption[63], ABCB1 for drug intake, lactase for enzyme digestion, SLCO1A2 (OATP3) or SLC10A2 (ASBT) for bile acid absorption[64], AQP1, 7, or 8 for water transportation[65], SLC22A1 through SLC22A5 (OCTs and OCTNs) or SCNN1B for cation transportation[66]. Taken together, we postulate that the metabolism of this group had been reprogrammed to involve certain cancer-related glycolysis burst events[67], which skews the co-culture model toward cancer metabolism. Therefore, the performance of certain drugs, such as long-chain fatty acid drugs like omega-3 fatty acids, lipid-lowering drugs like Gemfibrozil/Bezafibrate, cholesterol-lowering drugs like statins/ezetimibe, glucose-lowering drugs like sulfonylureas, and other drugs targeting the missing receptors, might be inaccurately evaluated in drug screening or permeability/efficacy tests. Interestingly, previous study reported the existence of ASBT in Caco-2[68], but we did not find its transcriptome in the co-culture.

While the model exhibited more small intestine (SI) characteristics (i.e., active SI-specific genes with most colon-specific genes missing), all absorptive populations expressed high levels of colon-specific ion transporter SLC26A2 and vitamin transporter RBP4 were expressed in all absorptive populations, indicating the model also possesses colon-like features given both Caco-2 and HT29 were CRC-derived. In unison, the heterogeneous ISCLC, TALC, and ECLC groups constitute a complex developmental roadmap (Fig. 3D).

Undifferentiated HT29, Goblet cell-like cells, and immature enteroendocrine cell-like cells were present in the model

Undifferentiated HT29 cells, which are characterized by their rapid proliferation and ability to form a monolayer in culture without showing functional differentiation typical of mature enterocytes, such as the formation of microvilli and the expression of brush border enzymes. We found undifferentiated HT29 cells, secretory progenitor-like cells, goblet-like cells, and 2 groups of immature enteroendocrine-like cells present in the co-culture system (Fig. 4A). Previous reports have shown that the majority (~95%) of HT29 monoculture is in an undifferentiated state, and they were defined such that they do not express definitive sets of

markers of known human epithelial cell types[69,70]. Admittedly, the profile of undifferentiated HT29, if any exists in the co-culture, is likely to differ substantially from that of the monoculture due to communication with Caco-2 cell and Transwell geometry. Therefore, we assigned the identities to U-HT29 and D-HT29 because of their lack of epithelial markers and share the most similarity with the profile of monoculture HT29 (Fig. 1H, Fig. 4B).

Most HT29 and secretory-like cells express CD24, CD44, AXIN2, and BMP4. CD24 and CD44, like discussed previously, are cell surface markers that are often used to identify cancer stem cells, which have the ability to self-renew and differentiate into various types of cancer cells. BMP4 is a bone morphogenetic protein that is reported to restrict the stemness of LGR5+ ISCs[71] and enhance the differentiation[72]. Such a BMP4 expression evokes similarity to Paneth or mesenchymal stromal cells because the currently accepted gut epithelial cells turnover model proposes that a BMP4 gradient increasing from the crypt to the villus tip guides the differentiation of ISCs, and a high level of BMP4 is found in the lamina propria of the villus tip, co-localizing with LGR5[73]. However, the possibility of this population being stromal cells was undermined by the high expression of epithelial cell adhesion molecule (EPCAM). Interestingly, this population also highly expressed lysozyme (LYZ) and end-of-differentiation marker keratin 20 (KRT20)[74]. Lysozyme (LYZ), a well-studied antimicrobial agent, is transcribed in BEST4+ enterocytes, PCs, tuft cells, and follicle-associated epithelial cells, which seemingly supports the hypothesis that those cells being Paneth-like cells. Earlier studies also claimed HT29's capability of differentiating into PCs based on its expression of LYZ[75]. However, the lack of other Paneth markers DEFA4/5 or ITLN2 (Fig. S2C) undermined the possibility that this population being full-fledged PC-like cells. KRT20 has long been used to identify the differentiated cells in the primary epithelium[74], yet its role in CRC cells is controversial. Although it was found to be highly associated with well-differentiated CRCs, it was also evident to decrease as CRCs differentiate[76]. Furthermore, the RNA-seq of the monoculture undifferentiated HT29 provided by CCLE reveals a high expression of KRT20, which suggests that KRT20 might not work as a differentiated state indicator like in the gut epithelium context but an intrinsic property of cancer. SECPLC featured a high expression of fetal secretory marker HES6[23], secretory differentiation modulator SOX4[77], and microtubule regulator STMN1[78] (Fig. 4B-C). It is also noteworthy that in the colorectal cancer context, these markers are associated with increased cell proliferation and invasion. The GCLC strongly and uniquely expressed gel-forming mucin MUC2[79], a putative marker of the intestinal goblet cells and the building block of the mucus layer (Fig. 4B-C). It also highly expressed mucin MUC5AC[80], which is the main component of the gastric and respiratory tract mucus layer that protects the tissue from stimulations yet is absent in the intestine[81]. Besides, part of the DEGs in human goblet cells were also present in the GCLC, such as IgGFc-binding protein FCGBP, SLC12A2, trefoil factor TFF3, and SPDEF (Fig. 4D). Although GCLC expressed some of the most well studied functional goblet genes, most of its the DEGs were not active in human goblet cells (Fig. S4C), suggesting the fundamental difference between healthy and malicious cells. Some of them, such as EGLN3[82], ADAM8[83], CA9[83], AQP3[84], and SIRT2[85], are related to the hypoxia response, suggesting a hypoxia-related cancer metabolism in the GCLC. Specifically, CA9 is a well-known hypoxia-inducible gene that is commonly overexpressed in cancer cells[86]. Additional experiments, such as measuring oxygen levels in the cells would be needed to confirm if such a response is caused by the presence of hypoxia or the oncogenic signaling. Also, the intrinsic hypoxia genes expression means the Caco-2/HT29 mix might respond differently to oxygen levels or nutrient availability, so when using them as a gut epithelium mimic and co-culturing them with bacteria, their response might not be transferable to the real gut. Thus, it is essential to carefully consider the experimental

conditions and controls to ensure that any observed effects are due to the bacterial co-culture and not to the Caco-2/HT29 cells' intrinsic properties. When comparing with human GCs, we found that 1433 out of 1647 commonly expressed genes in GCLC were also prevalent in human GCs (Fig. 2G). Overall, the HT29-derived goblet cells were relative comparable to its healthy human epithelial cells' analog.

Both early EECLC populations expressed immature EEC markers PAX4[87] and INSM1[88] (Fig. 4C). EECLC-1 uniquely expressed GCG, UCN3, ETV1 (EEC subtype L-cells), SST (EEC subtype D-cells)[23], whereas EECLC-2 expressed PROX1 (EEC subtype L-cells) and a high-level of STMN1 (Fig. 4B). Most of the mature human EEC DEGs such as CHGA and CHGB were missing in both EECLC groups (Fig. 4F), yet some EECLC DEGs were expressed in human EECs (Fig. S4 D). While only 46 genes were uniquely expressed in either type of EECLC and not in human EECs, a significant number of genes (3083) present in human EECs were not universally found in EECLCs (Fig. 4G), which supported our inference that EECLCs in the co-culture system were immature (i.e., had a less variated expression profile than mature EECs). Taken together, EECLCs were found in co-culture, but they might not be a perfect in vitro model of mature EEC.

We further subdivided the undifferentiated HT29 population into early, intermediate, and late stages (also referred to as differentiating) and proposed a potential differentiation pathway from early U-HT29 to GCLC (Fig. 4H-I). This is based on the understanding that GCLC originates from HT29 and not from Caco-2 in a static culture[70]. Early U-HT29 uniquely and highly expressed oncogenic/metastasis genes CD9[89], S100A4, and CRABP2[90] (Fig. 4I). The level of classic stem markers LGR5 and SMOC2 was undetectable (Fig. S2C), yet it expressed ASCL2 and RARRES2, a gene transcribed mainly in colon ISCs and in colon goblet cells[56] (Fig. 4B, Fig. 4I). The D-HT29 expressed metastasis markers insulin-like growth factor-binding protein IGFBP2[91] and chemokine receptor CXCR4 (Fig. 4I, Fig. S2C). CXCR4 is often overexpressed in cancer cells and is known to play a role in cancer progression and metastasis[92]. Its signaling has been shown to promote cancer cell survival, migration, and invasion, as well as angiogenesis. In fact, the CXCR4 is known to be the marker of metastasis cancer stem cells[93] (MCSCs) associated with the progression of CRC and a variety of cancers[94]. Other than IGFBP2 and CXCR4, the D-HT29 population also starts expressing secretory markers FOXA2[95] and SLC12A2[96] (Fig. 4B), indicating its secretory fate commitment. The intermediate U-HT29 cells didn't exhibit any differentially expressed genes compared to early U-HT29 and D-HT29. However, their profile demonstrated a smooth transition between the two groups, which justifies our use of the term "intermediate" to describe their state.

Co-culture model's response to the E. coli LPS treatment

In the human epithelium, Lipopolysaccharide (LPS) triggers an immune response, like neutrophil activation, by binding to Toll-like receptors (TLRs) on epithelial cells[97]. TLR4 is primarily responsible for initiating the defense-related NF-κB signaling pathway in cells when it binds to the LPS of gram-negative bacteria[98]. Previous studies employ Caco-2/HT29 as models to study intestinal epithelial cells' response to bacterial LPS by first confirming the presence of TLRs, or specifically TLR4. The undifferentiated HT29 is evident to express TLR4[99,100], but that expression in Caco-2 is controversial. Some studies find the presence of TLR4 in Caco-2 using the western blot and report more intense TLR4 expression[100,101]. Contradictorily, other studies report little TLR4 gene expression in Caco-2, with no observed LPS response[102–104]. This inconsistency is possibly due to the utilization of different anti-

TLR4 and other technical differences during the cell culture. The bulk RNA-seq reads from CCLC support no TLR4 transcription in Caco-2.

In our co-culture system, neither TLR4 nor most other TLRs were detected in any sequenced and qualified cells under either condition (Fig. S5B). It is possible if 1) Caco-2, as reported, did not express TLR4, and 2) after the HT29 grew to confluence, its TLR4 expression dramatically decreased[102].

Despite the lack of TLRs, we sought to understand how the system would change upon LPS treatment, given that another LPS-binding receptor, the glycosylphosphatidylinositol (GPI)-anchored protein (CD14), was present in the system, albeit at a very low level[105] (Fig. S5C). Compared to the control, treated cells were less variate (Fig. 5A). The amount of early ECLC-6 (potentially HT29 derived cells), early EECLC-2 (potentially Caco-2 derived EECLC), ISCLC 2 (FGG+ ISCLC) was less than expected, whereas the Early ECLC-7 (DEFB1 high early ECLC) was more than expected (Fig. 5B). We then interrogated the differential expression after LPS treatment in ISCLC, TALC, and ECLC specifically, given that the purpose of the model is to gain insights of human intestinal epithelial cells (Fig. 5C). GCLC was excluded since statistical tests cannot perform well on a small-size sample. Overall, the transcription of tight junction genes TJP1, OCLN, and CLDN1/7 did not significantly differ between conditions (Fig. S5A). Inflammatory genes TNF-α, IL1β, and IFN-γ were not present/ did not change significantly upon the treatment (Fig. S5B). In ISCLC, genes like absorptive fate determination gene HES1 and fatty acid biosynthesis rate-limiting enzyme SCD were upregulated, while the fetal ISC marker FGG and TFF2/3 were downregulated (Fig. 5C(i)-D). In TALC, vitamin carrier retinol-binding protein RBP4 decreased, and CEACAM6, a receptor for adherent-invasive E. coli adhesion[106], was upregulated (Fig. 5C(ii), E). SCD was also increased by the treatment in ECLC, while vitamin carrier retinol-binding protein RBP2 decreased (Fig. 5C(iii), F). Specifically, pathways related to differentiation, immune responses such as neutrophil activation, and metabolism were altered in the absorptive-like cells upon LPS treatment (Fig. 5G).

Discussion

The co-culture of Caco-2/HT29 has been widely used to simulate human intestinal epithelium in experiments. In our study, we carefully examined the variations in the Transwell co-culture, assessed how closely it resembled a healthy adult's epithelium, and studied its response to E. coli LPS. We identified the ISC, TA, EC, GC, and EEC-like cells together with differentiating HT29 cells in the system based on the expression of canonical markers. Overall, the co-cultured cells showed less transcription variety and activities than the human epithelial cells, with a unique, intrinsic cancerous properties that confers the co-culture similarity towards the fetal gut given the common re-expression of embryonic genes in the colorectal adenocarcinoma. Earlier studies have compared Caco-2 and HT29 cells with the fetal colon, but due to previous limitations in transcriptomics and proteomics, only a limited number of genes were reported[107–109]. Another example of discrepancy is that enterocytes-like cells in the model did not fully express the major transporters in the small intestine or colon. However, to date, none of the "major transporters" has been proven indispensable for physiological intestinal fatty acid transport[110]. One possible explanation is that intestinal fatty acid absorption might be regulated by two or more genes so that one gene is inactivated, and the other could compensate for its function[110]. Therefore, although the transcriptome of some transporters is absent, substances might still be able to be transported. We also found no TLR4 and very marginal TLR2 expression in the co-culture. Although further experiments

targeting the protein are necessary to confirm the absence of TLRs, here, we suggest that Caco-2/HT29 might not be an ideal model for studying microbe-epithelial cells interactions.

The Caco-2 and HT29 cell lines have proven extremely useful in replicating the gut epithelium, offering researchers a valuable laboratory model for exploring gastrointestinal physiology[111]. The Caco-2 cells, exhibiting enterocyte-like properties, primarily contribute to the understanding of absorptive processes, while the HT29 cells, characterized by mucus-secreting goblet cells, elucidate the protective mucosal barrier functions[46]. This powerful synergistic approach has not only accelerated advancements in the study of nutrient absorption and drug delivery but has also facilitated breakthroughs in understanding host-microbe interactions when immune cell presents[112], the effects of cytokines and inflammatory mediators on barrier function and the expression of tight junction proteins[113], the adherence and invasion of pathogenic bacteria[114], and molecular mechanisms of colorectal cancer[115], including the role of genetic mutations and the effects of potential therapeutic agents[116].

Caco-2 and HT29 cell lines are widely recognized as useful lab models for studying gut lining. Still, traditional animal models are often used to give a comprehensive, lifelike system for studying gut biology and confirming results from lab models. It facilitates the study of complex interactions among host genetics, immune responses, the microbiome, and different organs [117,118]. However, the ethical considerations, costs, and potential species-specific differences limit its universal applicability. Alternative in vitro models, such as gut-on-a-chip, have emerged recently, presenting unique advantages and drawbacks. It offers a dynamic, microfluidic environment that incorporates fluid shear stress and mechanical strain, more accurately simulating the physiological conditions of the gut[119,120]. Additionally, this platform allows for co-cultures of multiple cell types, including immune cells and microbes, thereby fostering a more comprehensive understanding of the intestinal ecosystem[121]. Compared to the primary-derived experimental methods like this, Caco-2 and HT29 cells exhibit a limited gene expression profile, fetal-like characteristics, and inherent cancer properties that might deviate from the in vivo gut epithelium. Nevertheless, their ease of use and cost-effectiveness still make them attractive for preliminary research.

Initially, we had 2 control groups and 2 LPS-treated groups. One of the LPS-treated groups was discarded due to its less-than-desirable sequencing quality. While this scope limits the scope of our analysis, we believe that our remaining treatment group, in combination with our controls, still provided valuable insights into our research question. We attempted to determine the origin of Caco-2/HT29 cells by analyzing the expression signatures we discovered from a comprehensive RNA sequence of Caco-2/HT29. However, we have to admit that during 22 days of co-culture, if the influences that two cell types exerted on each other were significant, then it is possible that cells developed convergently and acquired the pattern of the other type of cells so that their similarity with the mono-cultured cells is not indicative of their origins. Even so, as this study mainly focuses on describing the tissue similar to the epithelium formed after differentiation rather than the process of differentiation, we think this would be an interesting topic for future study.

Method

Germ-free coculture *in-vitro* Transwell gut model

The intestinal epithelial Transwell system was established by human colon carcinoma cell lines Caco-2 (passage no. 8) and HT29 (passage no. 13) (ATTC, United States) at $1x10^5$ cells/ cm^2 on permeable polyester membrane Transwell inserts. The cells were cultured in DMEM medium with Glutamax (Gibco, United States), supplemented with 10% (v/v) FBS,

and incubated in a humidified atmosphere (5% CO2, 37 °C) for 21 days to reach the full differentiation stage. Induction of barrier dysfunction in Caco-2 and HT-29 cell monolayers was induced in the apical chamber by 1 ng/ml *Escherichia coli O111:B4* lipopolysaccharides (LPS) stimulation on day 22 for 24 hours. On day 23, three Transwells were prepared for library construction following the manufacturer's instruction. The platform used for sequencing was Illumina NovaSeq 6000 S4.

Tight junction evaluation by transepithelial electrical resistance and the number of cells alive.

To determine tight junction integrity, transepithelial electrical resistance (TEER) was measured by an epithelial volt-ohmmeter (EVOM2; WPI, Berlin, Germany) with STX2 Chopstick probes. Measurements were performed at each culture medium exchange according to the manufacturer's instructions. The final TEER value was determined by subtracting the TEER measurement from a blank transwell insert and multiplying it by the cell culture surface.

Single Cell RNA seq Data Pre-processing

The FASTQ files of the sequencing data were input into the Cell Ranger[122] pipeline by 10x genomics to be filtered and mapped to the human genome GRCh38 to generate a count matrix that contains the number of detected transcriptomes of each gene for each cell. The downstream analysis was done with scanpy[123] (v1.9.1) on Python and Seurat[124–127] (v4.2.0) on R. Each dataset (control 1, control 2, and LPS-treated) was pre-processed separately. Low-quality cells were dropped based on their extremely low UMI counts, high mitochondrial gene counts, low ribosomal gene counts, and a low number of uniquely expressed genes (Table 1).

Table 1. Pre-processing Parameters

	Minimum counts	**Minimum genes**	**Maximum mitochondrial gene fraction**	**Minimum ribosomal gene fraction**
Control 1	2250	770		
Control 2	1500	700	0.2	0.05
LPS-treated	1500	620		

Genes expressed in less than 5 cells in each dataset were excluded. Multiplets were removed by Scrublet[128] (v0.2.3). After basic quality control, the 16019 genes' transcriptional profile in 13784 cells was obtained. The cell cycle score was calculated based on the expression of cell cycle-related genes as previously described[129]. Each dataset was normalized by SCTransform[130] (v0.3.5) on R. Cell's mitochondrial fraction and the difference between the S phase score and the G2M phase score, as proposed as the representation of the cell cycle score, were regressed during the normalization. Then, the top 3000 genes that were most variable in as many datasets as possible were used as anchors to integrate the datasets in

Seurat[131]. 3000 variable features were re-calculated for the combined dataset and used to perform the principal component analysis (PCA, 50 pcs). Leiden clustering[132] (resolution = 0.65) was performed based on the computed neighborhood graph of observations (UMAP [133], 50 pcs, size of neighborhood equals 15 cells) to reveal the general subtypes of the co-cultured cells. Partition-based graph abstraction[134] (PAGA) based on the Leiden clusters was used to initialize the uniform manifold approximation (UMAP, 50 pcs, min_dist = 0.1, spread = 1.1, n_components=2, alpha=1.0, gamma=1.0) to facilitate the convergence of manifold.

Subclustering and cell identity labeling

The logarithmized SCTransform corrected counts were used to perform the Wilcoxon rank-sum test across every Leiden cluster to find the signature of a cluster, which were genes that have log2 fold change > 3 and adjusted p-value < 0.01. The resultant signatures, together with the epithelial markers reported previously, confirm the general identity of clusters. Each cluster was further subclustered to reveal more heterogeneity among cells. The resolution used in each subclustering is listed (Table 2). The Wilcoxon rank-sum test was re-run on the finer clustered data to unveil the marker of each subgroup, with the same threshold for log2 fold change and adjusted p-value.

Table 2. Leiden Subclustering Resolution

	ISCLC	**TALC**	**Secretory**	**U-HT29**
Subclustering Resolution	0.2	0.25	0.3	0.35

Caco-2/HT29 Similarity Determination

The raw bulk mRNA-seq counts of HT29 and Caco-2 cells were obtained from Cancer Cell Line Encyclopedia. Reads were rounded to integers. The positive and negative count differences between the two cell lines were fitted into two negative binomial distributions using statsmodel[135] (v0.13.2), respectively. The overlap between the DEGs of each Leiden cluster and Caco-2/HT29 signatures were reported in Jaccard indices, normalized by the reference signatures size, and visualized in the heatmap.

Differentially Expressed Genes Analysis

To determine genes that were differentially expressed in the same type of cells between control and LPS-treated, Monocle3[136–139] (v1.2.9) was used to fit the SCTransform-corrected counts with a quasi-Poisson model to regress the gene expression over the treated/untreated condition. DE analysis was performed on treated/untreated ISCLC, TALC, and ECLC, given their adequate size for conducting statistical tests and their overall expression similarity to their human epithelial cells equivalent discussed earlier. The DE was defined as |log2 fold change| >= 0.5 with a Q value <= 0.05.

Data Availability

The Caco-2/HT29 coculture's scRNA-seq data can be accessed on NCBI GEO under GSE233628. Script used for analysis can be found on GitHub (https://github.com/WeldonSchool-BrubakerLab/Caco2-HT29-scRNAseq).

Authors Contribution

Conceptualization: J.M., L.G., D.B.; Methodology: R.R., J.M., S.J., D.B.; Analysis: R.R.; Writing–original draft: R.R., J.M.; Writing–review & editing: R.R., J.M., S.J., L.G., D.B.; Administration: D.B.

Competing Interest Statement

The authors have declared no competing interests.

Figure Captions

Fig. 1: The Caco-2/HT29 co-culture model produces highly heterogeneous cell populations resembling intestinal epithelial lineage.

A) Schematic of co-culture. B, C, D) uniform manifold approximation and projection (UMAP) of cell populations in the co-culture model with a resolution of (B) low (C) mid, and (D) high. ISCLC, intestinal stem-like cells; TALC, transit-amplifying-like cells; ECLC, enterocyte-like cells; SECPLC, secretory progenitor cell-like cells; GCLC, goblet cell-like cells; EECLC, enteroendocrine cell-like cells; U-HT29, undifferentiated HT29; D-HT29, differentiating HT29; FLC, fetal-like cells. E) UMAP visualization of co-cultured cells colored by key human intestinal epithelial lineage markers (cell types characterized by the markers are annotated next to the marker names) and by conditions (two controls, one LPS-treated group). F) heatmap of features in each cell population that have comparability with human intestinal epithelial lineages. G) composition of the co-cultured cells. H) conjectured Caco-2/HT29 origin of cell populations. Color bar) Jaccard similarity coefficient. I) Partition-based graph abstraction (PAGA) of the cell populations. Line thickness represents connectivity strength.

Fig. 2: Characterization of ISCLC/TALC

A) UMAP of the ISCLC/TALC colored by detailed cell subtypes. B) Dot plot showing the featured gene expression in each ISCLC/TALC subtype. C) UMAP of the ISCLC/TALC colored by key human ISC/TA markers. D, F) Heatmap showing the expression of human (D) ISC markers in ISCLC and (F) TA markers in TALC. E), G) Venn diagram showing the overlap between commonly expressed genes (>30% of the cells) in E) ISCLC and human ISC G) TALC and human TA.

Fig. 3: Characterization of ECLC

A) UMAP of the ECLC colored by detailed cell subtypes. B) UMAP of the ECLC colored by key human ECLC markers. C) Dot plot showing the expression of key transporters in each ELCL subtype. Sig, signatures; Chol, cholesterol. D) PAGA of the ECLC population. Line thickness represents connectivity strength.

Fig. 4: Characterization of secretory cell-like cells

A) UMAP of the HT29/secretory-like cells colored by detailed cell subtypes. B) Dot plot showing the featured gene expression in each HT29/secretory-like cells subtype. C) UMAP of the HT29/secretory-like cells colored by key human secretory (goblets/enteroendocrine) markers. D, F) Heatmap showing the expression of human (D) goblet markers in GCLC and (F) EEC markers in EECLC. E),G) Venn diagram showing the overlap between commonly expressed genes (>30% of the cells) in (E) GCLC and human goblet cells (G) EECLC and human EEC. H. UMAP of the HT29/secretory-like cells colored by (i) pseudotime, red = early, blue = late (ii) cell subtypes. I) Heatmap of gene expression changed along with the pseudotime. Expressions are standardized for each gene.

Fig. 5: LPS treatment results in mild yet statistically significant expression change in ISCLC, TALC, and ECLC

A) UMAP of all the co-cultured cells, colored in the density of LPS-treated cells. B) Stacked bar plots showing the fraction of LPS-treated cells in each detailed cell population. Assuming the LPS treatment does not alter the composition of the system, the expected fraction taken up by the treated cells in each population should equal the total percentage of treated cells (indicated by the dashed horizontal line). C) volcano plots of differentially expressed genes in (i) ISCLC (ii) TALC (iii) ECLC across two conditions. FC, fold-change. Orange dashed horizontal line, -log Q > 2. Orange dashed vertical lines, log2 FC > 0.5 or < -0.5. Part of the DEs is annotated in red on the graph. D, E, F) Violin plots of one gene upregulated and one gene downregulated by LPS treatment in (D) ISCLC (E) TALC (F) ECLC. G) Gene set enrichment analysis (GSEA) result of pathways (i) upregulated (ii) downregulated in ISCLC, (iii) upregulated (iv) downregulated in TALC, (v) upregulated (vi) downregulated in ECLC.

Fig. S1: Physical properties evaluation of the co-culture.
A) Brightfield microscopy shows the Caco-2 and HT29 monoculture morphology before co-culture. B) Brightfield microscopy shows the change in morphology and confluency during the co-culture. C) Trans-epithelial/endothelial electrical resistance (TEER) measurement with/without LPS treatment. Student t-test $p = 0.12$. D) Cell viability with/without LPS treatment. Student t-test $p = 0.15$.

Fig. S2: Quality control and supplementary expression maps
A) Violin plot of the mitochondrial fraction of each detailedly annotated cluster. B) Violin plot of the sequencing depth (i.e., counts of mRNA per cell) of each detailedly annotated cluster. C) Expression of various cell signatures on UMAP embedding.

Fig. S3: Expression maps of WNTs and CD14
A) Expression of WNTs and CD14 on UMAP embedding.

Fig. S4: Heatmaps of model DEGs' expression in corresponding human epithelial cells.
(A)ISCLC DEGs in human ISCs. (B) TALC DEGs in human Tas. (C)GCLC DEGs in human GCs. (D)EECLC DEGs in human EECs.

Fig. S5: Comparison of gene expressions between control and treated conditions. A) Violin plots of 4 representative intestinal epithelial cells tight junction genes expression across conditions. B) Violin plots of 3 inflammatory genes and TLRs expressions across conditions. C) Violin plot of detected CD14, TLR4, and LY96 transcription in the co-culture.

References

1. Ng, S. C. *et al.* Worldwide incidence and prevalence of inflammatory bowel disease in the 21st century: a systematic review of population-based studies. *Lancet* **390**, 2769–2778 (2017).
2. Baumgart, D. C. & Sandborn, W. J. Crohn's disease. *The Lancet* **380**, 1590–1605 (2012).
3. Dahlhamer, J. M., Zammitti, E. P., Ward, B. W., Wheaton, A. G. & Croft, J. B. Prevalence of Inflammatory Bowel Disease Among Adults Aged ≥18 Years — United States, 2015. *MMWR Morb Mortal Wkly Rep* **65**, 1166–1169 (2019).
4. Gisbert, J. P. & Chaparro, M. Systematic review with meta-analysis: inflammatory bowel disease in the elderly. *Aliment Pharmacol Ther* **39**, 459–477 (2014).
5. Jess, T., Rungoe, C. & Peyrin-Biroulet, L. Risk of colorectal cancer in patients with ulcerative colitis: a meta-analysis of population-based cohort studies. *Clin Gastroenterol Hepatol* **10**, 639–645 (2012).
6. Sambuy, Y. *et al.* The Caco-2 cell line as a model of the intestinal barrier: influence of cell and culture-related factors on Caco-2 cell functional characteristics. *Cell Biol Toxicol* **21**, 1–26 (2005).
7. Hubatsch, I., Ragnarsson, E. G. E. & Artursson, P. Determination of drug permeability and prediction of drug absorption in Caco-2 monolayers. *Nature Protocols 2007 2:9* **2**, 2111–2119 (2007).
8. Lesuffleur, T., Barbat, A., Dussaulx, E. & Zweibaum, A. Growth Adaptation to Methotrexate of HT-29 Human Colon Carcinoma Cells Is Associated with Their Ability to Differentiate into Columnar Absorptive and Mucus-secreting Cells1. *Cancer Res* **50**, 6334–6343 (1990).
9. Johansson, M. E. V. & Hansson, G. C. Immunological aspects of intestinal mucus and mucins. *Nature Reviews Immunology 2016 16:10* **16**, 639–649 (2016).
10. Shah, P., Jogani, V., Bagchi, T. & Misra, A. Role of Caco-2 cell monolayers in prediction of intestinal drug absorption. *Biotechnol Prog* **22**, 186–198 (2006).
11. Kolodziejczyk, A. A., Kim, J. K., Svensson, V., Marioni, J. C. & Teichmann, S. A. The technology and biology of single-cell RNA sequencing. *Mol Cell* **58**, 610–620 (2015).
12. Liu, Y. *et al.* Advances in Single-Cell Sequencing Technology and Its Application in Poultry Science. *Genes (Basel)* **13**, (2022).
13. Hoffmann, P. *et al.* Caco-2/HT29-MTX co-cultured cells as a model for studying physiological properties and toxin-induced effects on intestinal cells. *PLoS One* **16**, (2021).
14. Smith, N. R., Gallagher, A. C. & Wong, M. H. Defining a stem cell hierarchy in the intestine: markers, caveats and controversies. *J Physiol* **594**, 4781 (2016).

15. Sanman, L. E. *et al.* Transit-Amplifying Cells Coordinate Changes in Intestinal Epithelial Cell-Type Composition. *Dev Cell* **56**, 356-365.e9 (2021).
16. Chougule, P. *et al.* Isolation and characterization of human primary enterocytes from small intestine using a novel method. *Scand J Gastroenterol* **47**, 1334 (2012).
17. Nyström, E. E. L. *et al.* An intercrypt subpopulation of goblet cells is essential for colonic mucus barrier function. *Science* **372**, (2021).
18. Gehart, H. *et al.* Identification of Enteroendocrine Regulators by Real-Time Single-Cell Differentiation Mapping. *Cell* **176**, 1158-1173.e16 (2019).
19. Spit, M., Koo, B. K. & Maurice, M. M. Tales from the crypt: intestinal niche signals in tissue renewal, plasticity and cancer. *Open Biol* **8**, (2018).
20. Umar, S. Intestinal Stem Cells. *Curr Gastroenterol Rep* **12**, 340 (2010).
21. Lapidot, T. *et al.* A cell initiating human acute myeloid leukaemia after transplantation into SCID mice. *Nature* **367**, 645–648 (1994).
22. Elmentaite, R. *et al.* Cells of the human intestinal tract mapped across space and time. *Nature 2021 597:7875* **597**, 250–255 (2021).
23. Fawkner-Corbett, D. *et al.* Spatiotemporal analysis of human intestinal development at single-cell resolution. *Cell* **184**, 810-826.e23 (2021).
24. Elmentaite, R. *et al.* Single-Cell Sequencing of Developing Human Gut Reveals Transcriptional Links to Childhood Crohn's Disease. *Dev Cell* **55**, 771-783.e5 (2020).
25. Mounier, C. M. *et al.* Distinct expression pattern of the full set of secreted phospholipases A2 in human colorectal adenocarcinomas: sPLA2-III as a biomarker candidate. *Br J Cancer* **98**, 587–595 (2008).
26. Schewe, M. *et al.* Secreted Phospholipases A2 Are Intestinal Stem Cell Niche Factors with Distinct Roles in Homeostasis, Inflammation, and Cancer. *Cell Stem Cell* **19**, 38–51 (2016).
27. Jones, C., Avino, M., Giroux, V. & Boudreau, F. HNF4α Acts as Upstream Functional Regulator of Intestinal Wnt3 and Paneth Cell Fate. *CMGH* **15**, 593–612 (2023).
28. He, G. W. *et al.* Optimized human intestinal organoid model reveals interleukin-22-dependency of Paneth cell formation. *Cell Stem Cell* **29**, 1333-1345.e6 (2022).
29. Schmitt, M. *et al.* Paneth Cells Respond to Inflammation and Contribute to Tissue Regeneration by Acquiring Stem-like Features through SCF/c-Kit Signaling. *Cell Rep* **24**, 2312-2328.e7 (2018).
30. Li, N. *et al.* Heterogeneity in readouts of canonical Wnt pathway activity within intestinal crypts. *Dev Dyn* **245**, 822 (2016).
31. Shakiba, N. *et al.* CD24 tracks divergent pluripotent states in mouse and human cells. *Nat Commun* **6**, (2015).
32. Novak, D. *et al.* SOX2 in development and cancer biology. *Semin Cancer Biol* **67**, 74–82 (2020).
33. Wang, R. N. *et al.* Bone Morphogenetic Protein (BMP) signaling in development and human diseases. *Genes Dis* **1**, 87 (2014).
34. Mah, A. T., Yan, K. S. & Kuo, C. J. Wnt pathway regulation of intestinal stem cells. *J Physiol* **594**, 4837 (2016).
35. Fedi, P. *et al.* Isolation and biochemical characterization of the human Dkk-1 homologue, a novel inhibitor of mammalian Wnt signaling. *Journal of Biological Chemistry* **274**, 19465–19472 (1999).

36. Li, Q. *et al.* Lats1/2 Sustain Intestinal Stem Cells and Wnt Activation through TEAD-Dependent and Independent Transcription. *Cell Stem Cell* **26**, 675-692.e8 (2020).
37. Alrefaei, A. F. Frizzled receptors (FZD) play multiple cellular roles in development, in diseases, and as potential therapeutic targets. *J King Saud Univ Sci* **33**, 101613 (2021).
38. Hrckulak, D. *et al.* Wnt Effector TCF4 Is Dispensable for Wnt Signaling in Human Cancer Cells. *Genes (Basel)* **9**, (2018).
39. Rennoll, S. & Yochum, G. Regulation of MYC gene expression by aberrant Wnt/β-catenin signaling in colorectal cancer. *World J Biol Chem* **6**, 290 (2015).
40. Sun, Y. *et al.* Inhibitor of DNA binding 1 (Id1) mediates stemness of colorectal cancer cells through the Id1-c-Myc-PLAC8 axis via the Wnt/β-catenin and Shh signaling pathways. *Cancer Manag Res* **11**, 6855 (2019).
41. Kay, S. K. *et al.* The role of the Hes1 crosstalk hub in Notch-Wnt interactions of the intestinal crypt. *PLoS Comput Biol* **13**, (2017).
42. Topol, L., Chen, W., Song, H., Day, T. F. & Yang, Y. Sox9 Inhibits Wnt Signaling by Promoting β-Catenin Phosphorylation in the Nucleus. *J Biol Chem* **284**, 3323 (2009).
43. Rees, W. D., Tandun, R., Yau, E., Zachos, N. C. & Steiner, T. S. Regenerative Intestinal Stem Cells Induced by Acute and Chronic Injury: The Saving Grace of the Epithelium? *Front Cell Dev Biol* **8**, (2020).
44. Pandur, P., Läsche, M., Eisenberg, L. M. & Kühl, M. Wnt-11 activation of a non-canonical Wnt signalling pathway is required for cardiogenesis. *Nature* **418**, 636–641 (2002).
45. Yang, J. *et al.* Adenomatous polyposis coli (APC) differentially regulates beta-catenin phosphorylation and ubiquitination in colon cancer cells. *J Biol Chem* **281**, 17751–17757 (2006).
46. El-Bahrawy, M., Poulsom, S. R., Rowan, A. J., Tomlinson, I. T. & Alison, M. R. Characterization of the E-cadherin/catenin complex in colorectal carcinoma cell lines. *Int J Exp Pathol* **85**, 65 (2004).
47. Działo, E., Tkacz, K. & Błyszczuk, P. Crosstalk between the TGF-β and WNT signalling pathways during cardiac fibrogenesis. *Acta Biochim Pol* **65**, 341–349 (2018).
48. Kaiser, S. *et al.* Transcriptional recapitulation and subversion of embryonic colon development by mouse colon tumor models and human colon cancer. *Genome Biol* **8**, (2007).
49. White, R. B., Eisen, J. A., Kasten, T. L. & Fernald, R. D. Second gene for gonadotropin-releasing hormone in humans. *Proc Natl Acad Sci U S A* **95**, 305–309 (1998).
50. Gründker, C. & Emons, G. Role of gonadotropin-releasing hormone (GnRH) in ovarian cancer. *Reprod Biol Endocrinol* **1**, 65 (2003).
51. ZNF90 - Zinc finger protein 90 - Homo sapiens (Human) | UniProtKB | UniProt. https://www.uniprot.org/uniprotkb/Q03938/entry#function.
52. Finkbeiner, S. R. *et al.* Transcriptome-wide Analysis Reveals Hallmarks of Human Intestine Development and Maturation In Vitro and In Vivo. *Stem Cell Reports* **4**, 1140 (2015).
53. Miron, N. & Cristea, V. Enterocytes: active cells in tolerance to food and microbial antigens in the gut. *Clin Exp Immunol* **167**, 405 (2012).

54. Sun, H., Chow, E. C. Y., Liu, S., Du, Y. & Pang, K. S. The Caco-2 cell monolayer: usefulness and limitations. *http://dx.doi.org/10.1517/17425255.4.4.395* **4**, 395–411 (2008).
55. Haber, A. L. *et al.* A single-cell survey of the small intestinal epithelium. *Nature* **551**, 333 (2017).
56. Burclaff, J. *et al.* A Proximal-to-Distal Survey of Healthy Adult Human Small Intestine and Colon Epithelium by Single-Cell Transcriptomics. *Cell Mol Gastroenterol Hepatol* **13**, 1554–1589 (2022).
57. Simpson, I. A. *et al.* The facilitative glucose transporter GLUT3: 20 years of distinction. *Am J Physiol Endocrinol Metab* **295**, E242 (2008).
58. Kuo, C. C. *et al.* Metastatic Colorectal Cancer Rewrites Metabolic Program Through a Glut3-YAP-dependent Signaling Circuit. *Theranostics* **9**, 2526 (2019).
59. Song, M. Y., Lee, D. Y., Yun, S. M. & Kim, E. H. GLUT3 Promotes Epithelial–Mesenchymal Transition via TGF-β/JNK/ATF2 Signaling Pathway in Colorectal Cancer Cells. *Biomedicines* **10**, (2022).
60. Stahl, A. *et al.* Identification of the major intestinal fatty acid transport protein. *Mol Cell* **4**, 299–308 (1999).
61. Nassir, F., Wilson, B., Han, X., Gross, R. W. & Abumrad, N. A. CD36 is important for fatty acid and cholesterol uptake by the proximal but not distal intestine. *Journal of Biological Chemistry* **282**, 19493–19501 (2007).
62. Brown, J. M. & Yu, L. Opposing Gatekeepers of Apical Sterol Transport: Niemann-Pick C1-Like 1 (NPC1L1) and ATP-Binding Cassette Transporters G5 and G8 (ABCG5/ABCG8). *Immunol Endocr Metab Agents Med Chem* **9**, 18 (2009).
63. Koepsell, H. Glucose transporters in the small intestine in health and disease. *Pflugers Archiv* **472**, 1207 (2020).
64. Meier, P. J. & Stieger, B. Bile salt transporters. *Annu Rev Physiol* **64**, 635–661 (2002).
65. Wang, Y. *et al.* Single-cell transcriptome analysis reveals differential nutrient absorption functions in human intestine. *J Exp Med* **217**, (2020).
66. Tamai, I. Pharmacological and pathophysiological roles of carnitine/organic cation transporters (OCTNs: SLC22A4, SLC22A5 and Slc22a21). *Biopharm Drug Dispos* **34**, 29–44 (2013).
67. Wu, D. *et al.* Single-cell metabolic imaging reveals a SLC2A3-dependent glycolytic burst in motile endothelial cells. *Nature Metabolism 2021 3:5* **3**, 714–727 (2021).
68. van der Mark, V. A. *et al.* The lipid flippase heterodimer ATP8B1–CDC50A is essential for surface expression of the apical sodium-dependent bile acid transporter (SLC10A2/ASBT) in intestinal Caco-2 cells. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease* **1842**, 2378–2386 (2014).
69. Zweibaum, A. Differentiation of Human Colon Cancer Cells. *Pharmaceutical Applications of Cell and Tissue Culture to Drug Transport* 27–37 (1991) doi:10.1007/978-1-4757-0286-6_3.
70. Huet, C., Sahuquillo-Merino, C., Coudrier, E. & Louvard, D. Absorptive and mucus-secreting subclones isolated from a multipotent intestinal cell line (HT-29) provide new models for cell polarity and terminal differentiation. *J Cell Biol* **105**, 345–357 (1987).

71. Qi, Z. *et al.* BMP restricts stemness of intestinal Lgr5+ stem cells by directly suppressing their signature genes. *Nature Communications 2017 8:1* **8**, 1–14 (2017).
72. Zhang, L. *et al.* BMP4 administration induces differentiation of CD133+ hepatic cancer stem cells blocking their contributions to hepatocellular carcinoma. *Cancer Res* **72**, 4276–4285 (2012).
73. Bahar Halpern, K. *et al.* Lgr5+ telocytes are a signaling source at the intestinal villus tip. *Nature Communications 2020 11:1* **11**, 1–12 (2020).
74. Chan, C. W. M. *et al.* Gastrointestinal differentiation marker Cytokeratin 20 is regulated by homeobox gene CDX1. *Proc Natl Acad Sci U S A* **106**, 1936–1941 (2009).
75. Park, J. H., Lee, J. M., Lee, E. J., Kim, D. J. & Hwang, W. B. Kynurenine promotes the goblet cell differentiation of HT-29 colon carcinoma cells by modulating Wnt, Notch and AhR signals. *Oncol Rep* **39**, 1930–1938 (2018).
76. Cross, H. S. *et al.* Vitamin D receptor and cytokeratin expression may be progression indicators in human colon cancer. *Anticancer Res* **16**, 2333–2337 (1996).
77. Triana, S. *et al.* Single-cell transcriptomics reveals immune response of intestinal cell types to viral infection. *Mol Syst Biol* **17**, e9833 (2021).
78. Belletti, B. & Baldassarre, G. Stathmin: a protein with many tasks. New biomarker and potential target in cancer. *Expert Opin Ther Targets* **15**, 1249–1266 (2011).
79. Johansson, M. E. V., Holmén Larsson, J. M. & Hansson, G. C. The two mucus layers of colon are organized by the MUC2 mucin, whereas the outer layer is a legislator of host-microbial interactions. *Proc Natl Acad Sci U S A* **108**, 4659–4665 (2011).
80. Zhang, C. T., He, K. C., Pan, F., Li, Y. & Wu, J. Prognostic value of Muc5AC in gastric cancer: A meta-analysis. *World Journal of Gastroenterology□: WJG* **21**, 10453 (2015).
81. Pelaseyed, T. *et al.* The mucus and mucins of the goblet cells and enterocytes provide the first defense line of the gastrointestinal tract and interact with the immune system. *Immunol Rev* **260**, 8 (2014).
82. Ivan, M. & Kaelin, W. G. The EGLN-HIF O2-Sensing System: Multiple Inputs and Feedbacks. *Mol Cell* **66**, 772–779 (2017).
83. van den Beucken, T. *et al.* Hypoxia-induced Expression of Carbonic Anhydrase 9 Is Dependent on the Unfolded Protein Response. *J Biol Chem* **284**, 24204 (2009).
84. Malale, K. *et al.* Hypoxia-Induced Aquaporin-3 Changes Hepatocellular Carcinoma Cell Sensitivity to Sorafenib by Activating the PI3K/Akt Signaling Pathway. *Cancer Manag Res* **12**, 4321 (2020).
85. Kaitsuka, T., Matsushita, M. & Matsushita, N. Regulation of Hypoxic Signaling and Oxidative Stress via the MicroRNA–SIRT2 Axis and Its Relationship with Aging-Related Diseases. *Cells* **10**, (2021).
86. Pastorekova, S. & Gillies, R. J. The role of carbonic anhydrase IX in cancer development: links to hypoxia, acidosis, and beyond. *Cancer Metastasis Rev* **38**, 65 (2019).
87. Gehart, H. *et al.* Identification of Enteroendocrine Regulators by Real-Time Single-Cell Differentiation Mapping. *Cell* **176**, 1158-1173.e16 (2019).

88. Miller, S. A. *et al.* LSD1 and aberrant DNA methylation mediate persistence of enteroendocrine progenitors that support BRAF mutant colorectal cancer. *Cancer Res* **81**, 3791 (2021).
89. Rappa, G., Green, T. M., Karbanová, J., Corbeil, D. & Lorico, A. Tetraspanin CD9 determines invasiveness and tumorigenicity of human breast cancer cells. *Oncotarget* **6**, 7970 (2015).
90. Wu, J. I., Lin, Y. P., Tseng, C. W., Chen, H. J. & Wang, L. H. Crabp2 Promotes Metastasis of Lung Cancer Cells via HuR and Integrin β1/FAK/ERK Signaling. *Scientific Reports 2019 9:1* **9**, 1–14 (2019).
91. Png, K. J., Halberg, N., Yoshida, M. & Tavazoie, S. F. A microRNA regulon that mediates endothelial recruitment and metastasis by cancer cells. *Nature 2011 481:7380* **481**, 190–194 (2011).
92. Bianchi, M. E. & Mezzapelle, R. The Chemokine Receptor CXCR4 in Cell Proliferation and Tissue Regeneration. *Front Immunol* **11**, 2109 (2020).
93. Liao, W. T., Ye, Y. P., Deng, Y. J., Bian, X. W. & Ding, Y. Q. Metastatic cancer stem cells: from the concept to therapeutics. *Am J Stem Cells* **3**, 46 (2014).
94. Ma, J. *et al.* Fibroblast-derived CXCL12 regulates PTEN expression and is associated with the proliferation and invasion of colon cancer cells via PI3k/Akt signaling. *Cell Communication and Signaling* **17**, 1–12 (2019).
95. Lantz, K. A. *et al.* Foxa2 regulates multiple pathways of insulin secretion. *Journal of Clinical Investigation* **114**, 512 (2004).
96. Randall, J., Thorne, T. & Delpire, E. Partial cloning and characterization of Slc12a2: The gene encoding the secretory Na+-K+-2Cl- cotransporter. *Am J Physiol Cell Physiol* **273**, (1997).
97. Kawasaki, T. & Kawai, T. Toll-like receptor signaling pathways. *Front Immunol* **5**, 461 (2014).
98. Albensi, B. C. What is nuclear factor kappa B (NF-κB) doing in and to the mitochondrion? *Front Cell Dev Biol* **7**, 154 (2019).
99. Böcker, U. *et al.* Responsiveness of intestinal epithelial cell lines to lipopolysaccharide is correlated with Toll-like receptor 4 but not Toll-like receptor 2 or CD14 expression. *Int J Colorectal Dis* **18**, 25–32 (2003).
100. Furrie, E., Macfarlane, S., Thomson, G. & Macfarlane, G. T. Toll-like receptors-2, -3 and -4 expression patterns on human colon and their regulation by mucosal-associated bacteria. *Immunology* **115**, 565–574 (2005).
101. Panaro, M. A., Carofiglio, V., Acquafredda, A., Cavallo, P. & Cianciulli, A. Anti-inflammatory effects of resveratrol occur via inhibition of lipopolysaccharide-induced NF-κB activation in Caco-2 and SW480 human colon cancer cells. *British Journal of Nutrition* **108**, 1623–1632 (2012).
102. Sang, K. L. *et al.* Cellular differentiation-induced attenuation of LPS response in HT-29 cells is related to the down-regulation of TLR4 expression. *Biochem Biophys Res Commun* **337**, 457–463 (2005).
103. Grouls, M., van der Zande, M., de Haan, L. & Bouwmeester, H. Responses of increasingly complex intestinal epithelium in vitro models to bacterial toll-like receptor agonists. *Toxicology in Vitro* **79**, 105280 (2022).
104. Böcker, U. *et al.* Responsiveness of intestinal epithelial cell lines to lipopolysaccharide is correlated with Toll-like receptor 4 but not Toll-like receptor 2 or CD14 expression. *Int J Colorectal Dis* **18**, 25–32 (2003).

105. Funda, D. P. *et al.* CD14 Is Expressed and Released as Soluble CD14 by Human Intestinal Epithelial Cells In Vitro: Lipopolysaccharide Activation of Epithelial Cells Revisited. *Infect Immun* **69**, 3772 (2001).
106. Barnich, N. *et al.* CEACAM6 acts as a receptor for adherent-invasive E. coli, supporting ileal mucosa colonization in Crohn disease. *J Clin Invest* **117**, 1566–1574 (2007).
107. Levy, E. *et al.* Caco-2 cells and human fetal colon: a comparative analysis of their lipid transport. *Biochim Biophys Acta* **1439**, 353–362 (1999).
108. Fukui, K. *et al.* A fetal intestinal-type alkaline phosphatase produced in Caco-2 cells. *Tumour Biol* **17**, 320–324 (1996).
109. Zweibaum, A. *et al.* Immunohistological evidence, obtained with monoclonal antibodies, of small intestinal brush border hydrolases in human colon cancers and foetal colons. *Int J Cancer* **34**, 591–598 (1984).
110. Wang, T. Y., Liu, M., Portincasa, P. & Wang, D. Q. H. New insights into the molecular mechanism of intestinal fatty acid absorption. *Eur J Clin Invest* **43**, 1203 (2013).
111. Sambuy, Y. *et al.* The Caco-2 cell line as a model of the intestinal barrier: influence of cell and culture-related factors on Caco-2 cell functional characteristics. *Cell Biol Toxicol* **21**, 1–26 (2005).
112. Kim, H. J., Li, H., Collins, J. J. & Ingber, D. E. Contributions of microbiome and mechanical deformation to intestinal bacterial overgrowth and inflammation in a human gut-on-a-chip. *Proc Natl Acad Sci U S A* **113**, E7–E15 (2016).
113. Chiba, H., Osanai, M., Murata, M., Kojima, T. & Sawada, N. Transmembrane proteins of tight junctions. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1778**, 588–600 (2008).
114. Lesuffleur, T., Barbat, A., Dussaulx, E. & Zweibaum, A. Growth adaptation to methotrexate of HT-29 human colon carcinoma cells is associated with their ability to differentiate into columnar absorptive and mucus-secreting cells. *Cancer Res* **50**, 6334–6343 (1990).
115. Blouin, J. M. *et al.* Butyrate elicits a metabolic switch in human colon cancer cells by targeting the pyruvate dehydrogenase complex. *Int J Cancer* **128**, 2591–2601 (2011).
116. Artursson, P. Epithelial transport of drugs in cell culture. I: A model for studying the passive diffusion of drugs over intestinal absorptive (Caco-2) cells. *J Pharm Sci* **79**, 476–482 (1990).
117. Nguyen, T. L. A., Vieira-Silva, S., Liston, A. & Raes, J. How informative is the mouse for human gut microbiota research? *DMM Disease Models and Mechanisms* **8**, 1–16 (2015).
118. Hugenholtz, F. & de Vos, W. M. Mouse models for human intestinal microbiota research: a critical evaluation. *Cell Mol Life Sci* **75**, 149–160 (2018).
119. Walsh, D. I. *et al.* Emulation of Colonic Oxygen Gradients in a Microdevice. *SLAS Technol* **23**, 164–171 (2018).
120. Villenave, R. *et al.* Human Gut-On-A-Chip Supports Polarized Infection of Coxsackie B1 Virus In Vitro. *PLoS One* **12**, (2017).
121. Kim, H. J., Huh, D., Hamilton, G. & Ingber, D. E. Human gut-on-a-chip inhabited by microbial flora that experiences intestinal peristalsis-like motions and flow. *Lab Chip* **12**, 2165–2174 (2012).
122. Zheng, G. X. Y. *et al.* Massively parallel digital transcriptional profiling of single cells. *Nature Communications 2017 8:1* **8**, 1–12 (2017).

123. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: Large-scale single-cell gene expression data analysis. *Genome Biol* **19**, 1–5 (2018).
124. Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of single-cell gene expression data. *Nat Biotechnol* **33**, 495–502 (2015).
125. Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* (2021) doi:10.1016/j.cell.2021.04.048.
126. Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888–1902 (2019).
127. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol* **36**, 411–420 (2018).
128. Wolock, S. L., Lopez, R. & Klein, A. M. Scrublet: Computational Identification of Cell Doublets in Single-Cell Transcriptomic Data. *Cell Syst* **8**, 281-291.e9 (2019).
129. Tirosh, I. *et al.* Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352**, 189–196 (2016).
130. Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol* **20**, 1–15 (2019).
131. Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902.e21 (2019).
132. Traag, V. A., Waltman, L. & van Eck, N. J. From Louvain to Leiden: guaranteeing well-connected communities. *Scientific Reports 2019 9:1* **9**, 1–12 (2019).
133. McInnes, L., Healy, J. & Melville, J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. (2018).
134. Wolf, F. A. *et al.* PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. *Genome Biol* **20**, 1–9 (2019).
135. Seabold, S. & Perktold, J. statsmodels: Econometric and statistical modeling with python. in *9th Python in Science Conference* (2010).
136. Qiu, X. *et al.* Reverse graph embedding resolves complex single-cell developmental trajectories. *BioRxiv* (2017).
137. Cao, J. *et al.* The dynamics and regulators of cell fate decisions are revealed by pseudo-temporal ordering of single cells. *Nature* (2019).
138. Qiu, X. *et al.* Single-cell mRNA quantification and differential analysis with Census. *Nat Methods* (2017).
139. Trapnell, C. *et al.* The dynamics and regulators of cell fate decisions are revealed by pseudo-temporal ordering of single cells. *Nat Biotechnol* (2014).

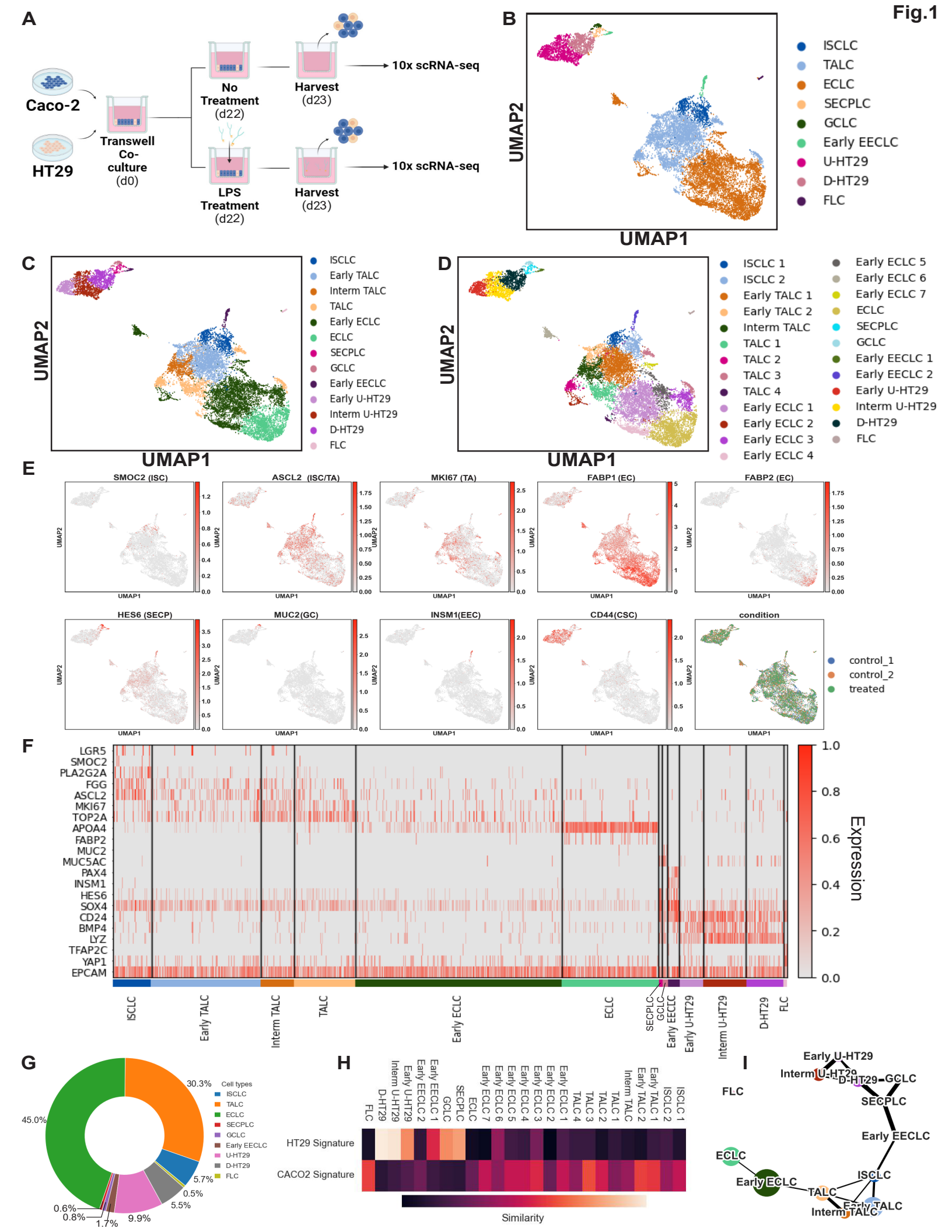


Fig.1
A
Caco-2
HT29
Transwell Co-culture (d0)
No Treatment (d22)
LPS Treatment (d22)
Harvest (d23)
Harvest (d23)
10x scRNA-seq
10x scRNA-seq
B
UMAP2
UMAP1
ISCLC
TALC
ECLC
SECPLC
GCLC
Early EECLC
U-HT29
D-HT29
FLC
C
UMAP2
UMAP1
ISCLC
Early TALC
Interm TALC
TALC
Early ECLC
ECLC
SECPLC
GCLC
Early EECLC
Early U-HT29
Interm U-HT29
D-HT29
FLC
D
UMAP2
UMAP1
ISCLC 1
ISCLC 2
Early TALC 1
Early TALC 2
Interm TALC
TALC 1
TALC 2
TALC 3
TALC 4
Early ECLC 1
Early ECLC 2
Early ECLC 3
Early ECLC 4
Early ECLC 5
Early ECLC 6
Early ECLC 7
ECLC
SECPLC
GCLC
Early EECLC 1
Early EECLC 2
Early U-HT29
Interm U-HT29
D-HT29
FLC
E
SMOC2 (ISC)
ASCL2 (ISC/TA)
MKI67 (TA)
FABP1 (EC)
FABP2 (EC)
HES6 (SECP)
MUC2(GC)
INSM1(EEC)
CD44(CSC)
condition
control_1
control_2
treated
F
LGR5
SMOC2
PLA2G2A
FGG
ASCL2
MKI67
TOP2A
APOA4
FABP2
MUC2
MUC5AC
PAX4
INSM1
HES6
SOX4
CD24
BMP4
LYZ
TFAP2C
YAP1
EPCAM
Expression
ISCLC
Early TALC
Interm TALC
TALC
Early ECLC
ECLC
SECPLC
GCLC
Early EECLC
Early U-HT29
Interm U-HT29
D-HT29
FLC
G
Cell types
ISCLC
TALC
ECLC
SECPLC
GCLC
Early EECLC
U-HT29
D-HT29
FLC
30.3%
5.7%
0.5%
5.5%
9.9%
1.7%
0.8%
0.6%
45.0%
H
HT29 Signature
CACO2 Signature
Similarity
I
Early U-HT29
Interm U-HT29
D-HT29
GCLC
SECPLC
Early EECLC
ECLC
Early ECLC
ISCLC
TALC
Early TALC
Interm TALC
FLC

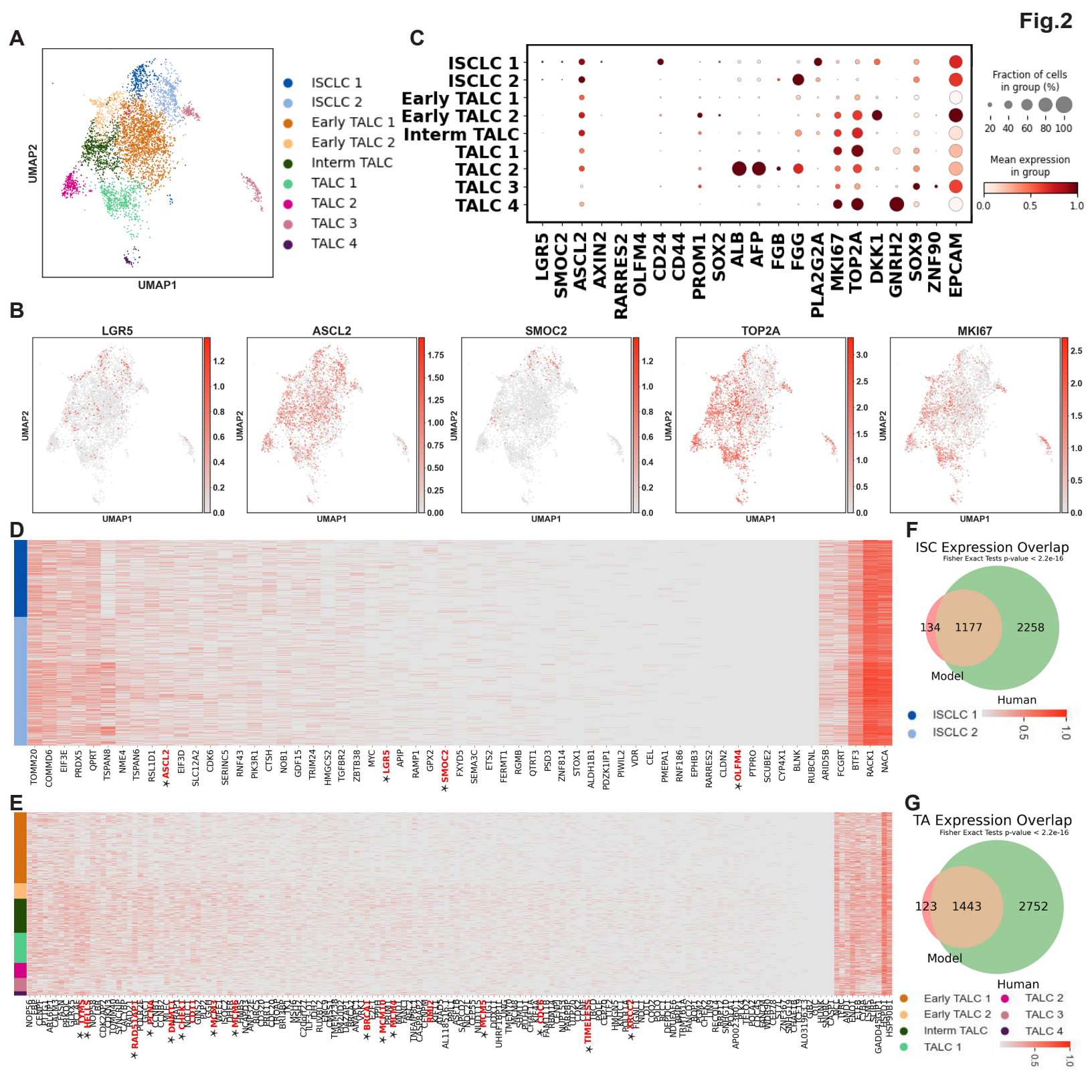
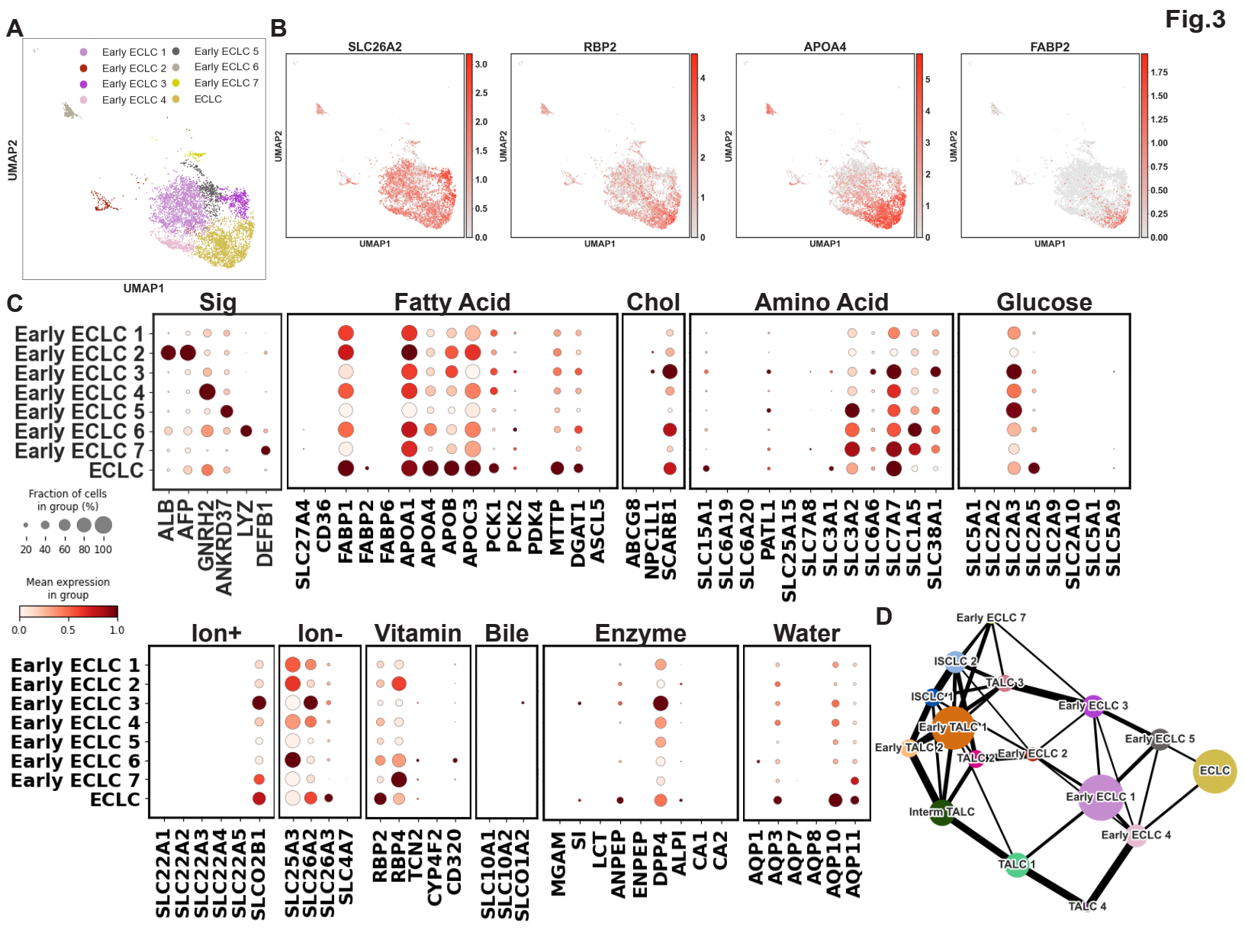


Fig.2
A
UMAP2
UMAP1
ISCLC 1
ISCLC 2
Early TALC 1
Early TALC 2
Interm TALC
TALC 1
TALC 2
TALC 3
TALC 4
C
ISCLC 1
ISCLC 2
Early TALC 1
Early TALC 2
Interm TALC
TALC 1
TALC 2
TALC 3
TALC 4
LGR5
SMOC2
ASCL2
AXIN2
RARRES2
OLFM4
CD24
CD44
PROM1
SOX2
ALB
AFP
FGB
FGG
PLA2G2A
MKI67
TOP2A
DKK1
GNRH2
SOX9
ZNF90
EPCAM
Fraction of cells in group (%)
20 40 60 80 100
Mean expression in group
0.0 0.5 1.0
B
LGR5
ASCL2
SMOC2
TOP2A
MKI67
UMAP2
UMAP1
D
TOMM20
COMMD6
EIF3E
PRDX5
QPRT
TSPAN8
NME4
TSPAN6
RSL1D1
ASCL2
EIF3D
SLC12A2
CDK6
SERINC5
RNF43
PIK3R1
CTSH
NOB1
GDF15
TRIM24
HMGCS2
TGFBR2
ZBTB38
MYC
LGR5
APIP
RAMP1
GPX2
SMOC2
FXYD5
SEMA3C
ETS2
FERMT1
RGMB
QTRT1
PSD3
ZNF814
STOX1
ALDH1B1
PDZK1IP1
PIWIL2
VDR
CEL
PMEPA1
RNF186
EPHB3
RARRES2
CLDN2
OLFM4
PTPRO
SCUBE2
CYP4X1
BLNK
RUBCNL
ARID5B
FCGRT
BTF3
RACK1
NACA
F
ISC Expression Overlap
Fisher Exact Tests p-value < 2.2e-16
134
1177
2258
Model
Human
ISCLC 1
ISCLC 2
0.5
1.0
E
G
TA Expression Overlap
Fisher Exact Tests p-value < 2.2e-16
123
1443
2752
Model
Human
Early TALC 1
Early TALC 2
Interm TALC
TALC 1
TALC 2
TALC 3
TALC 4
0.5
1.0

Fig.3


A
Early ECLC 1
Early ECLC 2
Early ECLC 3
Early ECLC 4
Early ECLC 5
Early ECLC 6
Early ECLC 7
ECLC
UMAP2
UMAP1
B
SLC26A2
RBP2
APOA4
FABP2
UMAP2
UMAP1
C
Sig
Fatty Acid
Chol
Amino Acid
Glucose
Early ECLC 1
Early ECLC 2
Early ECLC 3
Early ECLC 4
Early ECLC 5
Early ECLC 6
Early ECLC 7
ECLC
Fraction of cells in group (%)
20 40 60 80 100
Mean expression in group
0.0 0.5 1.0
ALB AFP GNRH2 ANKRD37 LYZ DEFB1
SLC27A4 CD36 FABP1 FABP2 FABP6 APOA1 APOA4 APOB APOC3 PCK1 PCK2 PDK4 MTTP DGAT1 ASCL5
ABCG8 NPC1L1 SCARB1
SLC15A1 SLC6A19 SLC6A20 PATL1 SLC25A15 SLC7A8 SLC3A1 SLC3A2 SLC6A6 SLC7A7 SLC1A5 SLC38A1
SLC5A1 SLC2A2 SLC2A3 SLC2A5 SLC2A9 SLC2A10 SLC5A1 SLC5A9
Ion+
Ion-
Vitamin
Bile
Enzyme
Water
SLC22A1 SLC22A2 SLC22A3 SLC22A4 SLC22A5 SLCO2B1
SLC25A3 SLC26A2 SLC26A3 SLC4A7
RBP2 RBP4 TCN2 CYP4F2 CD320
SLC10A1 SLC10A2 SLCO1A2
MGAM SI LCT ANPEP ENPEP DPP4 ALPI CA1 CA2
AQP1 AQP3 AQP7 AQP8 AQP10 AQP11
D
Early ECLC 7
ISCLC 2
TALC 3
ISCLC 1
Early ECLC 3
Early TALC 1
Early TALC 2
TALC 2
Early ECLC 2
Early ECLC 5
ECLC
Early ECLC 1
Interm TALC
Early ECLC 4
TALC 1
TALC 4

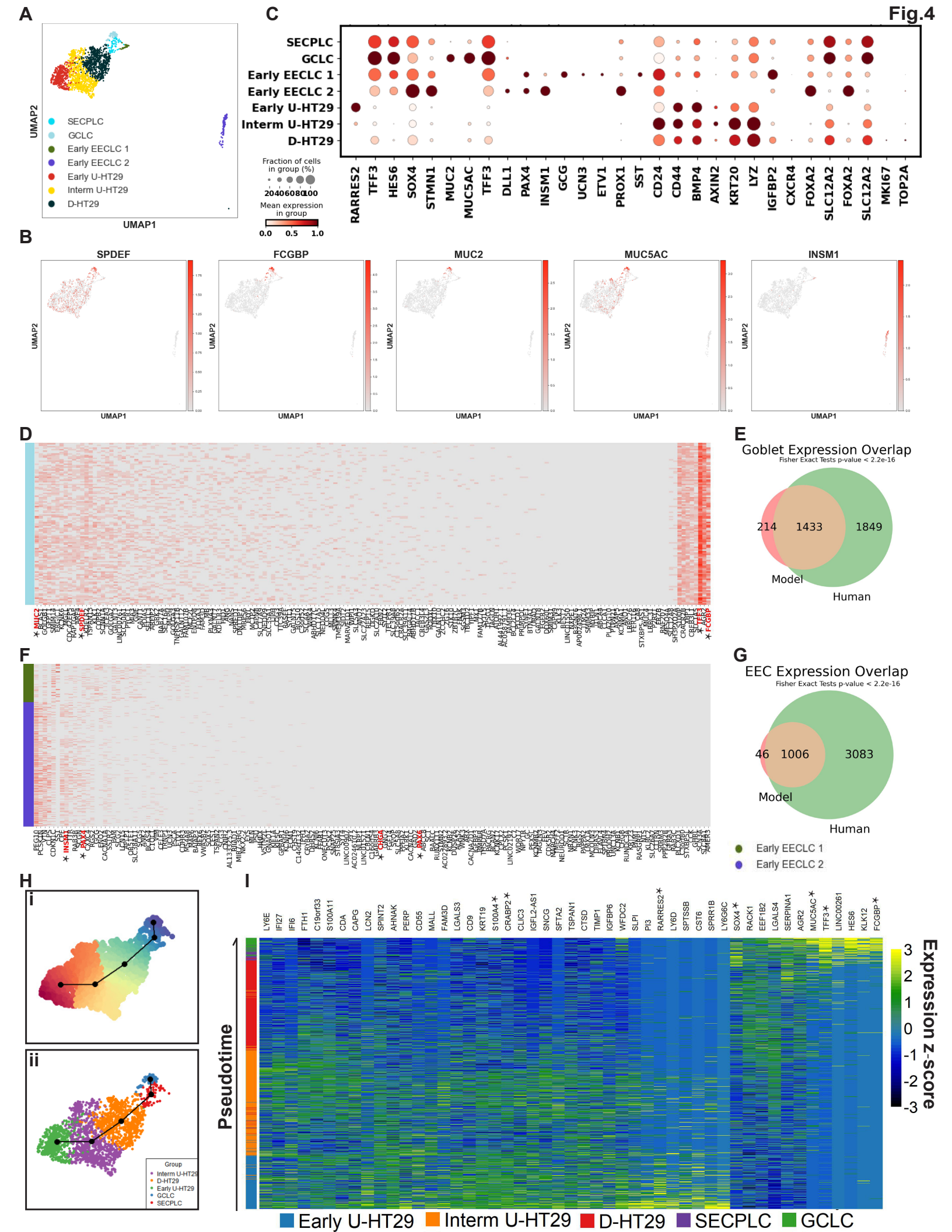


Fig.4
A
UMAP2
UMAP1
SECPLC
GCLC
Early EECLC 1
Early EECLC 2
Early U-HT29
Interm U-HT29
D-HT29
C
SECPLC
GCLC
Early EECLC 1
Early EECLC 2
Early U-HT29
Interm U-HT29
D-HT29
Fraction of cells in group (%)
20 40 60 80 100
Mean expression in group
0.0 0.5 1.0
RARRES2 TFF3 HES6 SOX4 STMN1 MUC2 MUC5AC TFF3 DLL1 PAX4 INSM1 GCG UCN3 ETV1 PROX1 SST CD24 CD44 BMP4 AXIN2 KRT20 LYZ IGFBP2 CXCR4 FOXA2 SLC12A2 FOXA2 SLC12A2 MKI67 TOP2A
B
SPDEF
FCGBP
MUC2
MUC5AC
INSM1
UMAP2
UMAP1
D
E
Goblet Expression Overlap
Fisher Exact Tests p-value < 2.2e-16
214
1433
1849
Model
Human
F
G
EEC Expression Overlap
Fisher Exact Tests p-value < 2.2e-16
46
1006
3083
Model
Human
Early EECLC 1
Early EECLC 2
H
i
ii
Group
Interm U-HT29
D-HT29
Early U-HT29
GCLC
SECPLC
I
Pseudotime
Expression z-score
3
2
1
0
-1
-2
-3
Early U-HT29
Interm U-HT29
D-HT29
SECPLC
GCLC

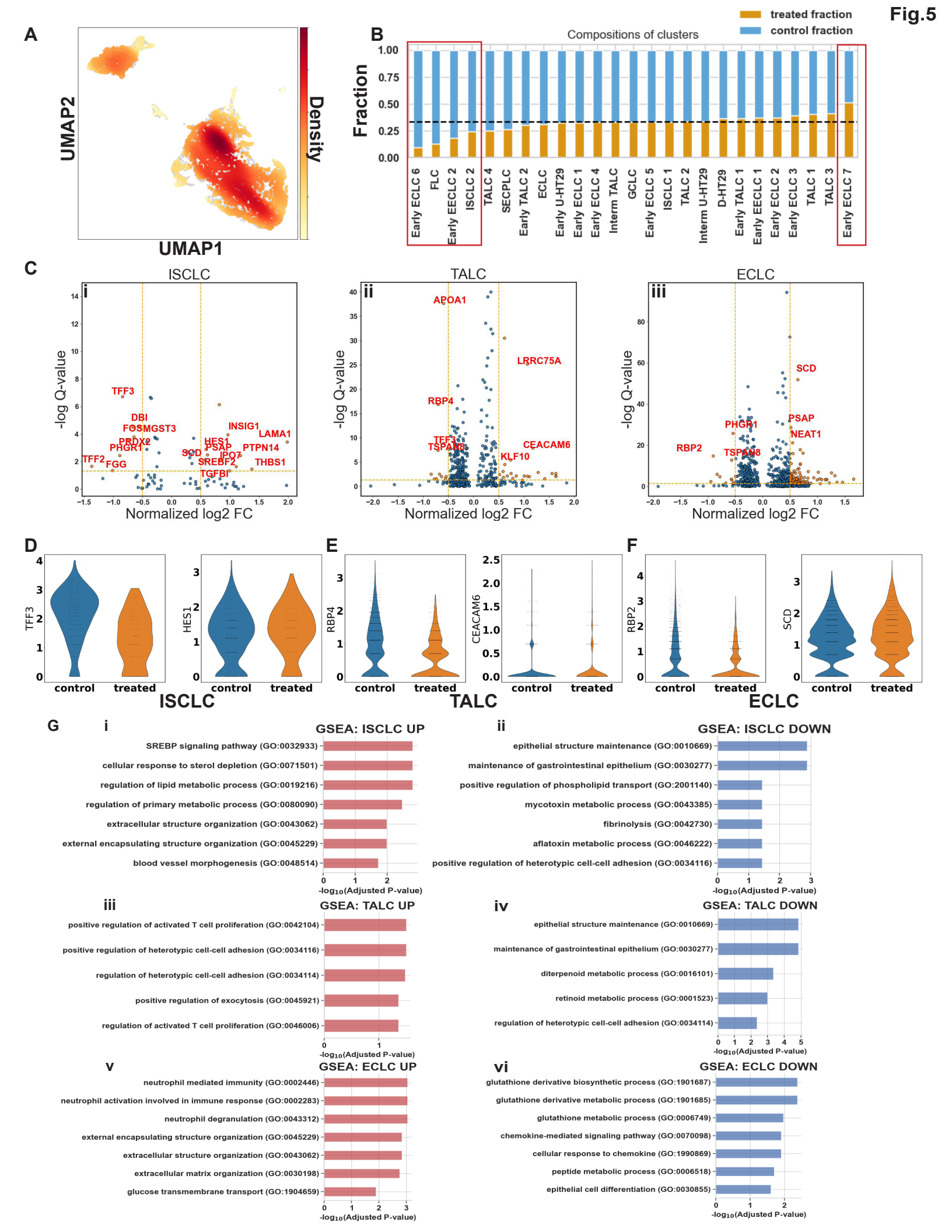
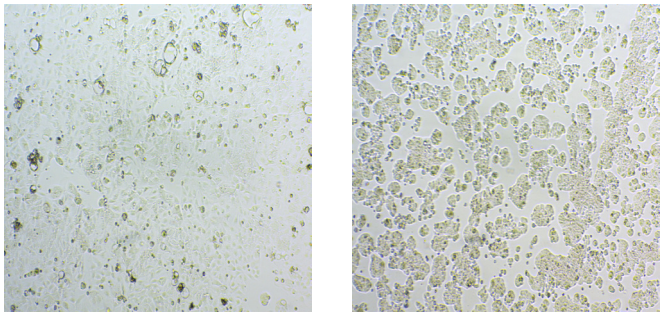


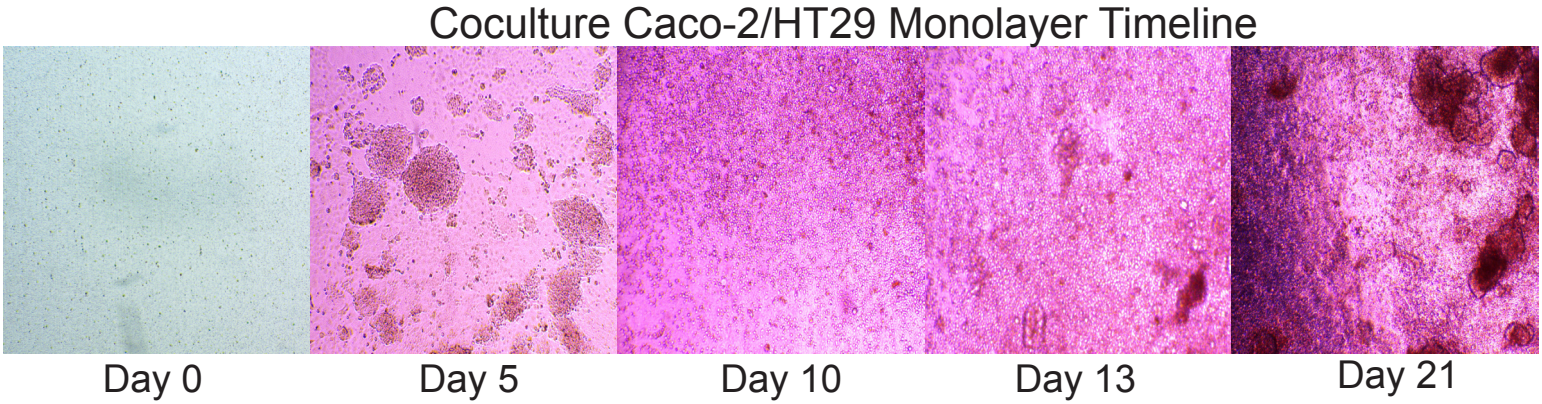
Fig.5
A
UMAP2
UMAP1
Density
B
Compositions of clusters
treated fraction
control fraction
Fraction
1.00
0.75
0.50
0.25
0.00
Early ECLC 6
FLC
Early EECLC 2
ISCLC 2
TALC 4
SECPLC
Early TALC 2
ECLC
Early U-HT29
Early ECLC 1
Early ECLC 4
Interm TALC
GCLC
Early ECLC 5
ISCLC 1
TALC 2
Interm U-HT29
D-HT29
Early TALC 1
Early EECLC 1
Early ECLC 2
Early ECLC 3
TALC 1
TALC 3
Early ECLC 7
C
ISCLC
TALC
ECLC
i
ii
iii
-log Q-value
Normalized log2 FC
TFF3
DBI
FOSMGST3
PRDX2
PHGR1
TFF2
FGG
INSIG1
LAMA1
HES1
PSAP
PTPN14
SCD
IPO7
SREBF2
THBS1
TGFBI
APOA1
LRRC75A
RBP4
TFF3
TSPAN8
CEACAM6
KLF10
SCD
PSAP
PHGR1
NEAT1
RBP2
TSPAN8
D
TFF3
HES1
control
treated
ISCLC
E
RBP4
CEACAM6
control
treated
TALC
F
RBP2
SCD
control
treated
ECLC
G
i
GSEA: ISCLC UP
SREBP signaling pathway (GO:0032933)
cellular response to sterol depletion (GO:0071501)
regulation of lipid metabolic process (GO:0019216)
regulation of primary metabolic process (GO:0080090)
extracellular structure organization (GO:0043062)
external encapsulating structure organization (GO:0045229)
blood vessel morphogenesis (GO:0048514)
-log10(Adjusted P-value)
ii
GSEA: ISCLC DOWN
epithelial structure maintenance (GO:0010669)
maintenance of gastrointestinal epithelium (GO:0030277)
positive regulation of phospholipid transport (GO:2001140)
mycotoxin metabolic process (GO:0043385)
fibrinolysis (GO:0042730)
aflatoxin metabolic process (GO:0046222)
positive regulation of heterotypic cell-cell adhesion (GO:0034116)
-log10(Adjusted P-value)
iii
GSEA: TALC UP
positive regulation of activated T cell proliferation (GO:0042104)
positive regulation of heterotypic cell-cell adhesion (GO:0034116)
regulation of heterotypic cell-cell adhesion (GO:0034114)
positive regulation of exocytosis (GO:0045921)
regulation of activated T cell proliferation (GO:0046006)
-log10(Adjusted P-value)
iv
GSEA: TALC DOWN
epithelial structure maintenance (GO:0010669)
maintenance of gastrointestinal epithelium (GO:0030277)
diterpenoid metabolic process (GO:0016101)
retinoid metabolic process (GO:0001523)
regulation of heterotypic cell-cell adhesion (GO:0034114)
-log10(Adjusted P-value)
v
GSEA: ECLC UP
neutrophil mediated immunity (GO:0002446)
neutrophil activation involved in immune response (GO:0002283)
neutrophil degranulation (GO:0043312)
external encapsulating structure organization (GO:0045229)
extracellular structure organization (GO:0043062)
extracellular matrix organization (GO:0030198)
glucose transmembrane transport (GO:1904659)
-log10(Adjusted P-value)
vi
GSEA: ECLC DOWN
glutathione derivative biosynthetic process (GO:1901687)
glutathione derivative metabolic process (GO:1901685)
glutathione metabolic process (GO:0006749)
chemokine-mediated signaling pathway (GO:0070098)
cellular response to chemokine (GO:1990869)
peptide metabolic process (GO:0006518)
epithelial cell differentiation (GO:0030855)
-log10(Adjusted P-value)

Fig.S1

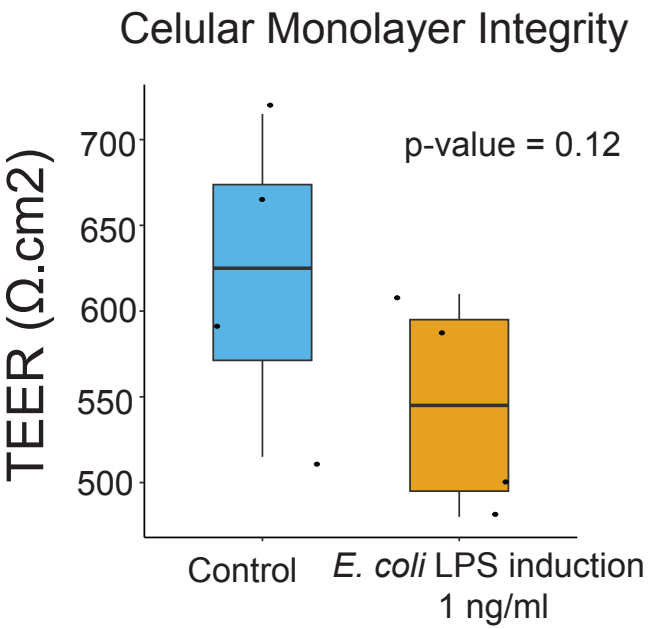
A


Intestinal Epithelial Transwell System
Caco-2
HT29

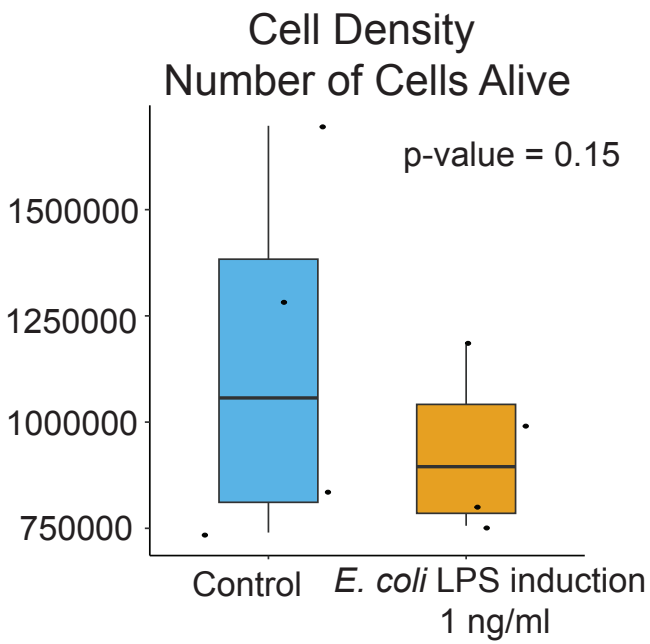
B


Coculture Caco-2/HT29 Monolayer Timeline
Day 0
Day 5
Day 10
Day 13
Day 21

C


Celular Monolayer Integrity
TEER (Ω.cm2)
700
650
600
550
500
p-value = 0.12
Control
E. coli LPS induction
1 ng/ml

D


Cell Density
Number of Cells Alive
1500000
1250000
1000000
750000
p-value = 0.15
Control
E. coli LPS induction
1 ng/ml

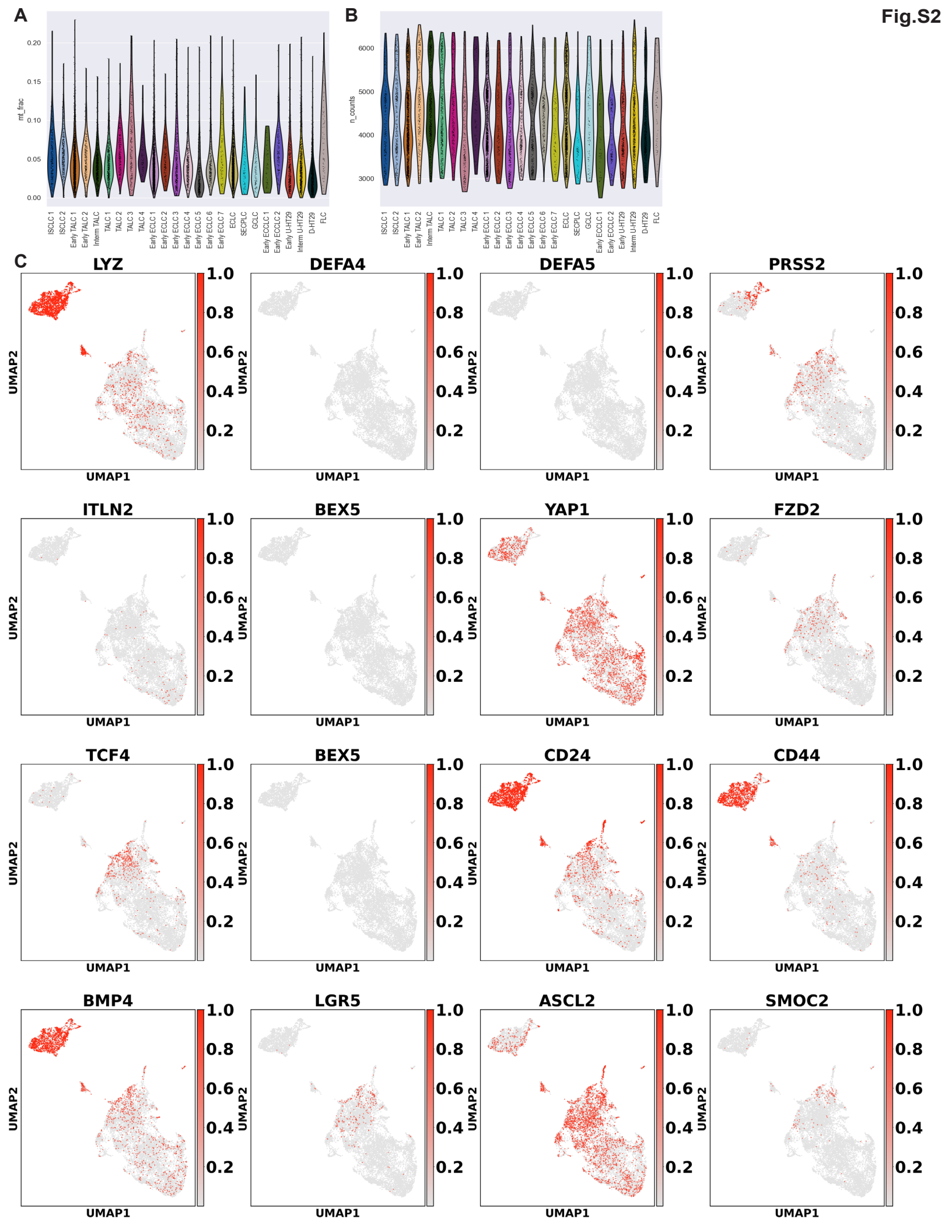
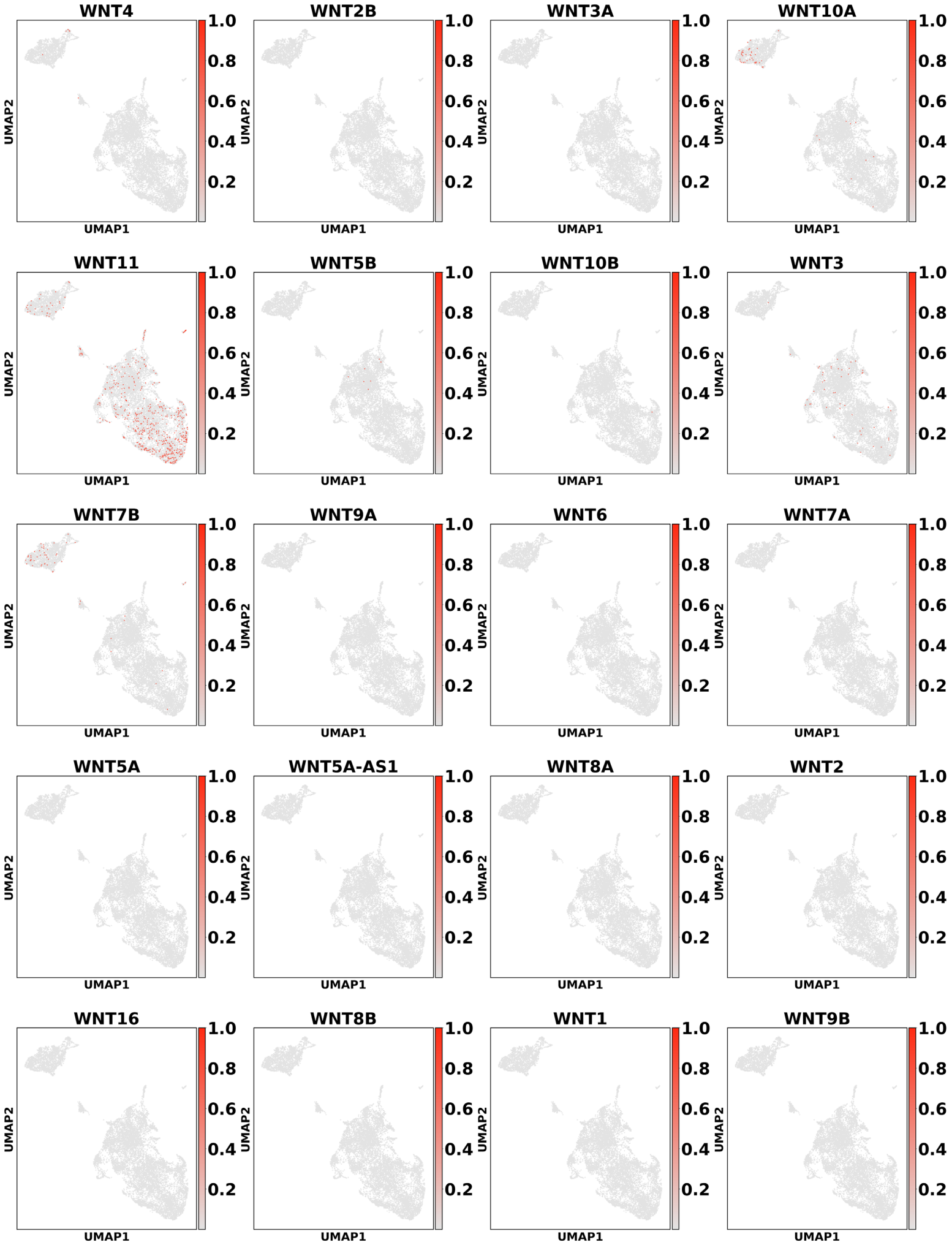


Fig.S2
A
mt_frac
0.20
0.15
0.10
0.05
0.00
B
n_counts
6000
5000
4000
3000
ISCLC 1
ISCLC 2
Early TALC 1
Early TALC 2
Interm TALC
TALC 1
TALC 2
TALC 3
TALC 4
Early ECLC 1
Early ECLC 2
Early ECLC 3
Early ECLC 4
Early ECLC 5
Early ECLC 6
Early ECLC 7
ECLC
SECPLC
GCLC
Early ECCLC 1
Early ECCLC 2
Early U-HT29
Interm U-HT29
D-HT29
FLC
C
LYZ
DEFA4
DEFA5
PRSS2
ITLN2
BEX5
YAP1
FZD2
TCF4
BEX5
CD24
CD44
BMP4
LGR5
ASCL2
SMOC2
UMAP1
UMAP2
1.0
0.8
0.6
0.4
0.2

A

Fig.S3


WNT4
WNT2B
WNT3A
WNT10A
WNT11
WNT5B
WNT10B
WNT3
WNT7B
WNT9A
WNT6
WNT7A
WNT5A
WNT5A-AS1
WNT8A
WNT2
WNT16
WNT8B
WNT1
WNT9B
UMAP1
UMAP2
1.0
0.8
0.6
0.4
0.2

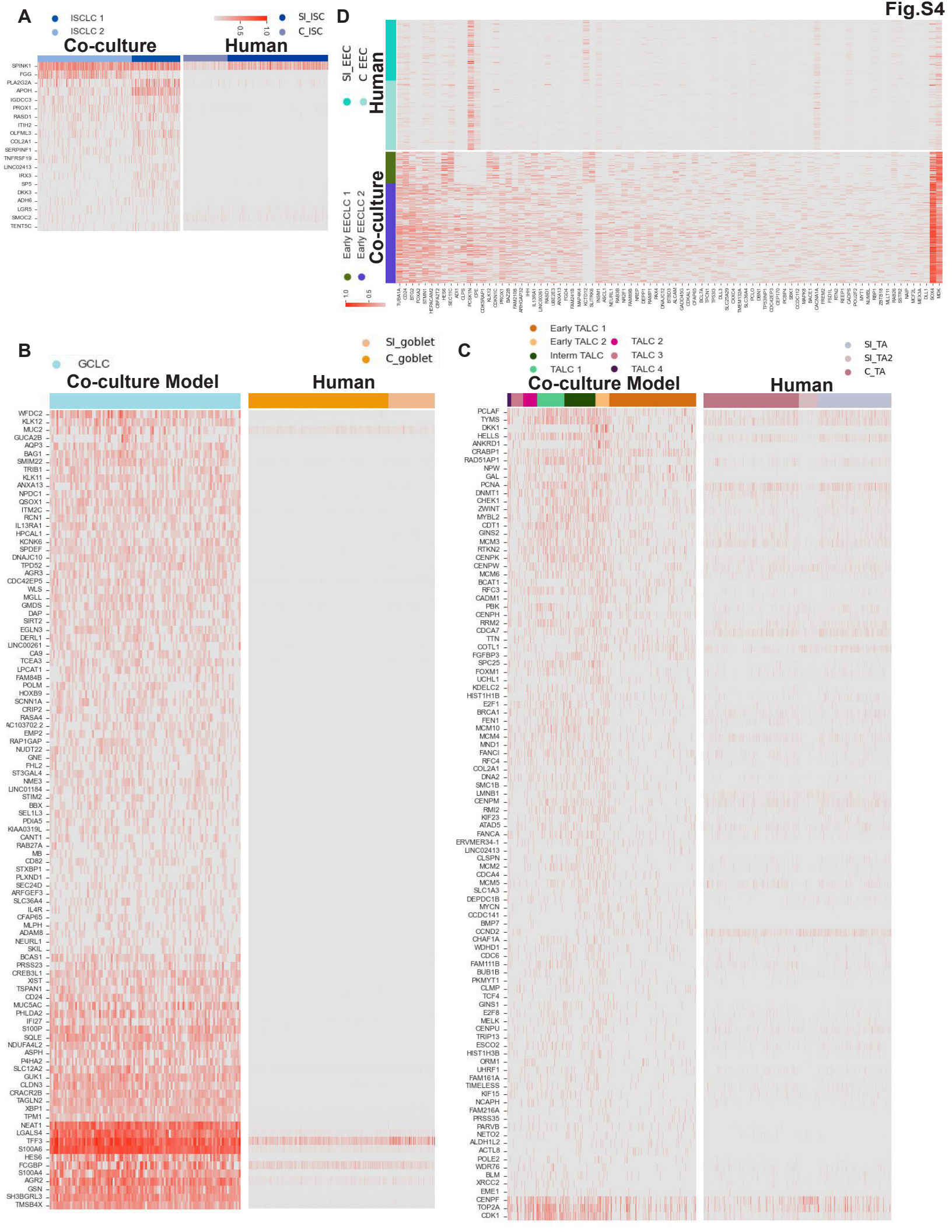
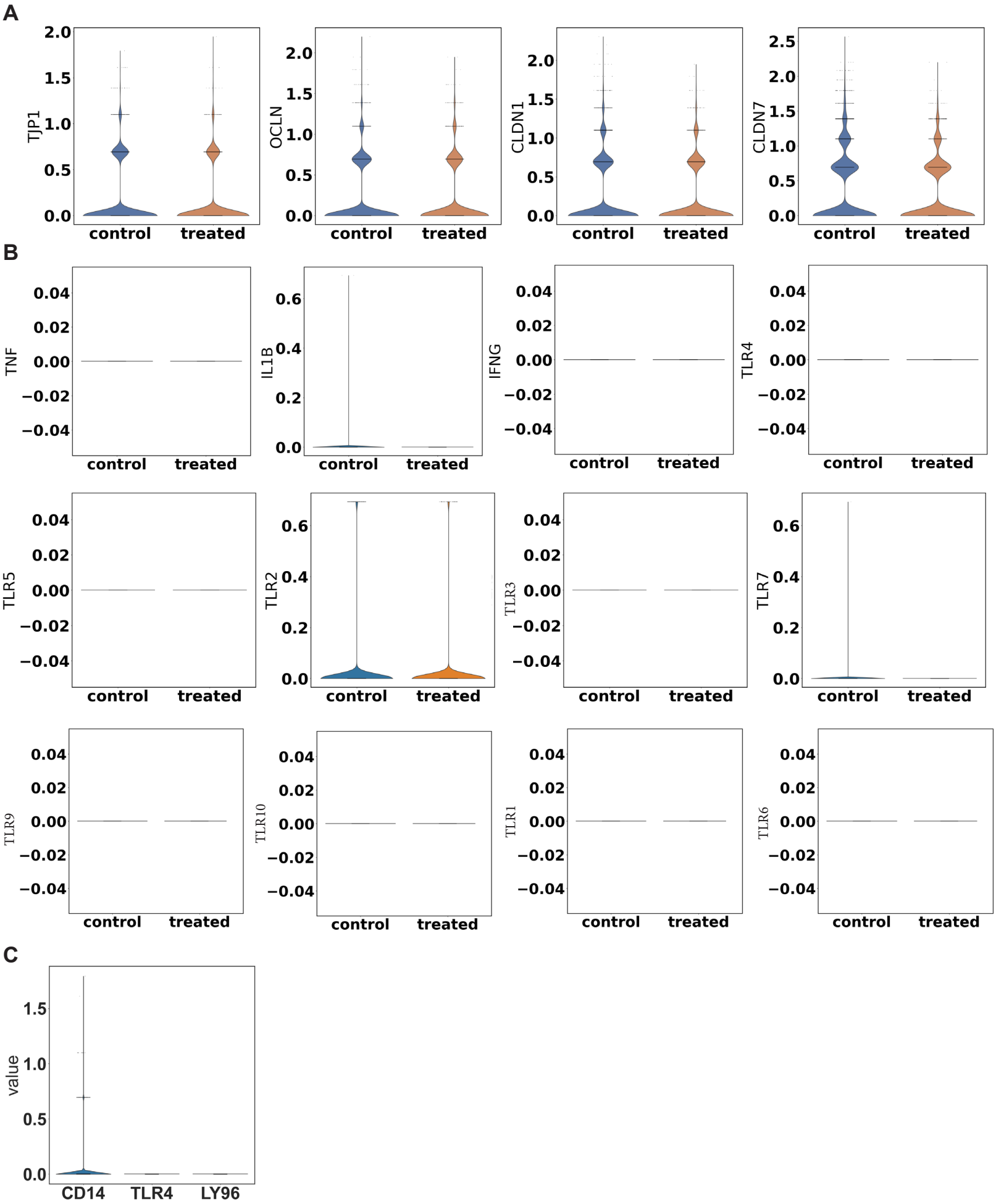


Fig.S4
A
ISCLC 1
ISCLC 2
Co-culture
Human
SI_ISC
C_ISC
SPINK1
FGG
PLA2G2A
APOH
IGDCCC3
PROX1
RASD1
ITIH2
OLFML3
COL2A1
SERPINF1
TNFRSF19
LINC02413
IRX3
SP5
DKK3
ADH6
LGR5
SMOC2
TENT5C
D
SI_EEC
C_EEC
Human
Early EECLC 1
Early EECLC 2
Co-culture
B
GCLC
Co-culture Model
Human
SI_goblet
C_goblet
WFDC2
KLK12
MUC2
GUCA2B
AQP3
BAG1
SMIM22
TRIB1
KLK11
ANXA13
NPDC1
QSOX1
ITM2C
RCN1
IL13RA1
HPCAL1
KCNK6
SPDEF
DNAJC10
TPD52
AGR3
CDC42EP5
WLS
MGLL
GMDS
DAP
SIRT2
EGLN3
DERL1
LINC00261
CA9
TCEA3
LPCAT1
FAM84B
POLM
HOXB9
SCNN1A
CRIP2
RASA4
AC103702.2
EMP2
RAP1GAP
NUDT22
GNE
FHL2
ST3GAL4
NME3
LINC01184
STIM2
BBX
SEL1L3
PDIA5
KIAA0319L
CANT1
RAB27A
MB
CD82
STXBP1
PLXND1
SEC24D
ARFGEF3
SLC36A4
IL4R
CFAP65
MLPH
ADAM8
NEURL1
SKIL
BCAS1
PRSS23
CREB3L1
XIST
TSPAN1
CD24
MUC5AC
PHLDA2
IFI27
S100P
SQLE
NDUFA4L2
ASPH
P4HA2
SLC12A2
GUK1
CLDN3
CRACR2B
TAGLN2
XBP1
TPM1
NEAT1
LGALS4
TFF3
S100A6
HES6
FCGBP
S100A4
AGR2
GSN
SH3BGRL3
TMSB4X
C
Early TALC 1
Early TALC 2
Interm TALC
TALC 1
TALC 2
TALC 3
TALC 4
Co-culture Model
Human
SI_TA
SI_TA2
C_TA
PCLAF
TYMS
DKK1
HELLS
ANKRD1
CRABP1
RAD51AP1
NPW
GAL
PCNA
DNMT1
CHEK1
ZWINT
MYBL2
CDT1
GINS2
MCM3
RTKN2
CENPK
CENPW
MCM6
BCAT1
RFC3
CADM1
PBK
CENPH
RRM2
CDCA7
TTN
COTL1
FGFBP3
SPC25
FOXM1
UCHL1
KDELC2
HIST1H1B
E2F1
BRCA1
FEN1
MCM10
MCM4
MND1
FANCI
RFC4
COL2A1
DNA2
SMC1B
LMNB1
CENPM
RMI2
KIF23
ATAD5
FANCA
ERVMER34-1
LINC02413
CLSPN
MCM2
CDCA4
MCM5
SLC1A3
DEPDC1B
MYCN
CCDC141
BMP7
CCND2
CHAF1A
WDHD1
CDC6
FAM111B
BUB1B
PKMYT1
CLMP
TCF4
GINS1
E2F8
MELK
CENPU
TRIP13
ESCO2
HIST1H3B
ORM1
UHRF1
FAM161A
TIMELESS
KIF15
NCAPH
FAM216A
PRSS35
PARVB
NETO2
ALDH1L2
ACTL8
POLE2
WDR76
BLM
XRCC2
EME1
CENPF
TOP2A
CDK1

Fig.S5


A
TJP1
OCLN
CLDN1
CLDN7
control
treated
B
TNF
IL1B
IFNG
TLR4
TLR5
TLR2
TLR3
TLR7
TLR9
TLR10
TLR1
TLR6
C
value
CD14
TLR4
LY96