Deterministic scRNA-seq captures variation in intestinal crypt and organoid composition

Johannes Bues[1,2,8], Marjan Biočanin[1,2,8], Joern Pezoldt[1,2,8], Riccardo Dainese[1,2], Antonius Chrisnandy[3], Saba Rezakhani[3], Wouter Saelens[4,5], Vincent Gardeux[1,2], Revant Gupta[2,6], Rita Sarkis[1], Julie Russeil[1,2], Yvan Saeys[4,5], Esther Amstad[7], Manfred Claassen[2,6], Matthias Lutolf[3], Bart Deplancke[1,2,*]

1 Laboratory of Systems Biology and Genetics, Institute of Bioengineering, School of Life Sciences, Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland

2 Swiss Institute of Bioinformatics (SIB), Lausanne, Switzerland

3 Laboratory for Stem Cell Bioengineering, Institute of Bioengineering, School of Life Sciences, Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland

4 Data mining and Modelling for Biomedicine, VIB Center for Inflammation Research, Ghent, Belgium

5 Department of Applied Mathematics, Computer Science and Statistics, Ghent University, Ghent, Belgium

6 Institute for Molecular Systems Biology, Eidgenössische Technische Hochschule Zürich (ETH Zürich), Zürich, Switzerland

7 Soft Materials Laboratory, Institute of Materials, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland

8 These authors contributed equally. Johannes Bues, Marjan Biočanin, Joern Pezoldt

* To whom correspondence should be addressed. Bart Deplancke: bart.deplancke@epfl.ch

Abstract

Single-cell RNA-sequencing (scRNA-seq) approaches have transformed our ability to resolve cellular properties across systems, but are currently tailored toward large cell inputs (> 1,000 cells). This renders them inefficient and costly when processing small, individual tissue samples, which tends to be resolved by loading bulk samples, yielding confounded mosaic cell population read-outs. Here, we developed a deterministic, mRNA-capture bead and cell co-encapsulation dropleting system, DisCo, aimed at processing low-input samples (<500 cells). We demonstrate that DisCo enables precise particle and cell positioning and droplet sorting control through combined machine-vision and multilayer microfluidics, enabling continuous processing of low-input single cell suspensions at high capture efficiency (> 70%) and speeds up to 350 cells per hour. To underscore DisCo's unique capabilities, we analyzed 31 individual intestinal organoids at varying developmental stages. This revealed extensive organoid heterogeneity, identifying distinct subtypes including i) a regenerative fetal-like *Ly6a*$^{+}$ stem cell population which persists as symmetrical cysts, or spheroids, even under differentiation conditions, and ii) a so far uncharacterized "gobloid" subtype consisting predominantly of precursor and mature (*Muc2*$^{+}$) goblet cells. To complement this dataset and to demonstrate DisCo's capacity to process low-input, *in vivo*-derived tissues, we also analyzed individual mouse intestinal crypts. This revealed the existence of crypts with compositional similarity to spheroids, i.e. predominantly consisting of regenerative stem cells, suggesting the existence of regenerating crypts in the homeostatic intestine. These findings demonstrate the unique power of DisCo in providing high-resolution snapshots of cellular heterogeneity among small, individual tissues.

Introduction

Single-cell RNA sequencing (scRNA-seq)[1] induced a paradigm shift in biomedical sciences, since it allows the dissection of cellular heterogeneity by high-dimensional data. Recent technological developments, particularly for cell capture and reaction compartmentalization[2–6], have led to a substantial increase in experimental throughput, enabling massive mapping efforts such as the mouse and human cell-atlas studies[5,7,8]. These developments were accompanied by biochemical advances, for instance for targeted transcript detection or sample multiplexing[9,10], which present a rich toolbox for large-scale scRNA-seq studies. However, since the majority of methods rely on stochastic cell capture, entailing large sample inputs, efficient processing of small samples (< 1,000 cells) remains challenging. The three main reasons for this are: 1) high fixed run costs, which lead to a large expense per cell at low inputs. For instance, a 10X Chromium run on 100 cells would cost $44 per sequenced cell. To reduce the cost per cell, cell hashing approaches have been developed (e.g. CITE-seq[10], MULTI-seq[11], or ClickTags[12]) that enable the parallelized processing of samples and that are clearly valuable to increase throughput and to reduce experimental costs as well as batch effects. However, unavoidable cell losses that occur during extensive cell washing and that are required to purge superfluous barcoding molecules render these approaches infeasible for the multiplexing of small input cell samples. 2) Requirements of minimum cell inputs. For example, index-sorting FACS or 10X Chromium require minimum cellular inputs ranging between 10,000 and 500 cells, respectively[13,14]. 3) Reduced effectiveness at low inputs because of limited cell capture efficiencies or cell size-selective biases[15] when processing small heterogeneous samples. To illustrate these limitations, we summarized the performance of various scRNA-seq technologies on low input samples in **Table 1**. Consequently, small samples, involving for instance zebrafish embryos[16], organisms like *C. elegans*[17], or intestinal organoids[18–20], are still pooled to obtain cell numbers that are compatible with stochastic microfluidic and well-based technologies. Thus, it is rather paradoxical that limitations overcome by single cell methods are nevertheless reintroduced at the sample level: artificial averages

across samples, resulting in an inability to resolve cell type distributions of individual systems or tissues. This particularly hampers research on emergent and self-organizing multicellular systems, such as organoids, that are heterogeneous and small at critical development stages.

In this study, we develop a novel deterministic, mRNA-capture bead and cell co-encapsulation dropleting system (DisCo) for low input scRNA-seq. In contrast to established methods that rely on passive cell capture strategies, we utilize machine-vision to actively detect cells and coordinate their capture in droplets. This active flow control approach allows for continuous operation, enabling free per run scaling and serial processing of samples. We demonstrate that DisCo can efficiently process samples of 100 cells and below, a sample type that tends to fall outside the scope of current cell processing platforms (**Table 1**). Indeed, the fully automated precision cell capture process makes this platform particularly well suited for the routine handling of small, individual tissues. Here, we exploit DisCo's unique capabilities to explore the heterogeneous, early development of individual intestinal organoids at the single cell level. Grown from single stem cells, organoids of vastly different morphologies and cell type compositions form under seemingly identical *in vitro* conditions[18]. These unpredictable developmental patterns represent one of the major limitations of this model system, preventing their widespread implementation e.g. in drug screens[21]. Thus, efforts to advance our understanding of the extent of organoid heterogeneity, how it arises, and how it can be controlled, for instance with synthetic growth matrices[22,23], are of essence. In depth mapping of individual organoid heterogeneity by scRNA-seq has so far been prevented by the minute cell numbers contained in a single intestinal organoid at critical developmental stages, such as post symmetry breaking at the 16-32 cell stage[18]. In total, we "DisCo'd" 31 single organoids at four developmental time points post symmetry breaking, and identified striking differences in cell type composition between individual organoids. Among these subtypes, we detected "spheroids" that are composed of regenerative fetal-like stem cells marked by Stem Cell Antigen-1 (*Sca1*/*Ly6a*)[24–27] and that persist under differentiation conditions.

In addition, we uncovered a rare subtype that is predominantly composed of precursor- and mature goblet cells and which we termed “gobloids”. Finally, we used DisCo to analyze individual, intestinal crypts, revealing variation in the cellular composition of crypts, while providing a proof-of-principle for our technology’s capacity to also process low-cell input, *in vivo*-derived tissues.

Results

To develop our Deterministic Co-encapsulation (DisCo) system, we engineered a three inlet (cells, beads, oil) multilayer dropleting device with two outlet ports (sample, waste) (Schematic **Figure 1A**, full design **Supplementary Figure 1A**). On this device, each inlet and outlet was augmented with a Quake-style microvalve[28] to facilitate flow control during operation. In addition, one common valve spanning both the cell and bead channel, termed the dropleting valve, was integrated to allow for on-demand droplet generation. To operate the device, we developed a three-stage process (**Figure 1B**): 1. Stop two particles at the encapsulation site, 2. Eject particles into one droplet, 3. Selectively extract the droplet in a sample channel (Microscopy images of the process are depicted in **Figure 1C**). To enable precise coordination of particles in microchannels, we developed a machine vision-based approach utilizing subsequent image subtraction for blob detection (**Supplementary Figure 1B**), and on-chip valves for flow-control. Deterministic displacement patterns were induced by opening and closing the cell and bead valves (depicted in **Supplementary Figure 1C)**, which moved particles according to discrete jumps into the target region of interest (ROI) with 95.9% of particles placed in an approximately ~200 µm-wide region (**Supplementary Figure 1D**). Upon placement, the stopped particles were ejected by pressurizing the dropleting valve, displacing an equal volume of liquid from both channels. The ejected liquid phase was then sheared into a droplet by activating the oil stream. We found that precise pressurization of the dropleting valve allowed for accurate control of droplet volume (**Supplementary Figure 1E**, **Supplementary Video 1**). Post droplet formation, the outlet valves were actuated to separate the formed droplet from the excess waste liquids (**Figure 1D**). With all

components operating in tight orchestration, we were able to generate monodisperse emulsions with high co-encapsulation purity (**Figure 1E**, **Supplementary Video 2**).

As a first benchmarking experiment, we set out to determine the encapsulation performance of DisCo for scRNA-seq-related applications, involving co-encapsulation of single cells with microspheres. Specifically, we aimed to reconfigure the Drop-seq[2] approach as it only requires coordination of two channels, as compared to three channels for inDrop[3]. Since co-encapsulation purity and cell capture efficiency are critical system parameters for droplet scRNA-seq systems, we quantified the system's processing speed and encapsulation performance in a free-run configuration, i.e. without cell number limitations at varying cell densities. We found that on average, 91.4% of all droplets contain a cell and a bead, and 1.7% contain an independent cell doublet (**Figure 1F**). Overall, the system provided high cell capture efficiencies of 90% at around 200 cells per hour for a 2 cells/µL cell concentration (**Figure 1G**). At higher cell concentrations of 20 cells/µL, the processing speed could be increased to 350 cells per hour, yet with decreased capture efficiencies of approximately 75%.

Next, we benchmarked the performance of DisCo for scRNA-seq. With drastically reduced bead amounts contained in the generated sample emulsion, we utilized our previously developed and characterized chip-based cDNA generation protocol[29]. Initially, as a library quality measure, we performed a species-mixing experiment of human HEK 293T and murine brown pre-adipocyte IBA cells. We observed clear species separation (**Figure 1H**), consistent with the limited number of previously detected doublets (**Figure 1F**), and increased read-utilization rate compared to conventional Drop-seq experiments (**Supplementary Figure 1F**). As previously reported[30], we found that our data displayed a skewed barcode sequence editing distance distribution compared to a true random distribution (**Supplementary Figure 1G**). Since the uniquely low number of beads in DisCo samples (< 500) renders the random occurrence of barcode sequences with an

editing distance < 3 rare, we developed a graph-based approach to identify and merge closely related barcodes (described in **Material and Methods**). We found that this approach did not compromise the single cell purity (**Supplementary Figure 1H**) and improved the detectable number of transcripts per cell as compared to published Drop-seq datasets on HEK 293T cells[2,29] (**Figure 1I**). Since DisCo requires longer time periods to process cells (e.g. compared to the 10X Chromium instrument), we also assessed time-dependent effects on the quality of the single cell data by analyzing HEK 293T cells that were loaded on our system (at room temperature) for 0 - 20, 20 - 40, and 40 - 60 minutes. Furthermore, we sampled cells that were stored for 120 and 180 minutes on ice. Cell stress metrics such as mitochondrial read content and heat shock protein expression revealed that loading / storage of the cells did not negatively affect the cells (**Supplementary Figure 1J**). However, further gene expression analysis by integration of all samples in one common dimensionality revealed that cells that had been stored for 180 minutes on ice exhibited a more demarcated distribution of cells in the map, suggesting that storage of cell suspensions for extended periods (> 2h) of time starts to introduce artifacts (**Supplementary Figure 1K**).

Since DisCo actively controls fluid flow on the microfluidic device, we observed that the system requires negligible run-in time, and is capable of efficiently processing cells from the first cell on. Given this observation, and the high-capture efficiency of DisCo in free-run mode, we hypothesized that the system should provide reliable performance on small samples of 100 cells and below. To determine the overall cell capture efficiency of DisCo, we precisely quantified the number of input cells using impedance measurements. Specifically, we utilized custom pipette tips augmented with a DISPENCELL gold-plated electrode, which allowed accurate counting of the number of input cells as validated by microscopy (**Supplementary Figure 1I**). Utilizing the DISPENCELL approach, we processed cell numbers between 50 - 200 cells, of which on average 86.4% (SD ± 8.1%) were visible on the chip. Of all input cells, 79.1% (SD ± 7.4%) were

157 successfully co-encapsulated, which corresponds to a co-encapsulation efficiency of 91.6% (SD
158 ± 1.6%) of all visible cells, while 74.9% (SD ± 10.7%) of input cells were found as barcodes over
159 500 UMIs per cell (**Figure 1J**). To contextualize these performance metrics, we performed similar
160 experiments on the Fluidigm C1 platform. This is because, based on reported data, this platform
161 appears the most efficient compared to other scRNA-seq technologies in processing low-input
162 cell samples (**Table 1**), rendering it thus DisCo's closest competitor. Specifically, we performed
163 three independent experiments utilizing 38, 125, and 215 HEK 293T cells, as quantified by
164 microscopy in order to be compatible with the C1 protocol. We chose the 96-trap chip in
165 combination with the SMART-seq v4 protocol, since, according to the user manual, it is the more
166 suitable chip for low cell inputs. We found that the Fluidigm C1 achieved absolute processing
167 efficiencies between 30 – 45% (**Supplementary Figure 1L**). Noteworthy, we were able to
168 reproduce the performance listed by the manufacturer for the 215-cell condition, suggesting that
169 our experiments were in good accordance with the expected efficiencies of the C1 system. We
170 believe that these values support the data reported in **Table 1**, and again emphasize the challenge
171 linked to processing low-input, single cell samples. Overall, these results, together with reported
172 data, indicate that the DisCo approach outperforms other technologies that are capable of
173 analyzing low-input cell samples in terms of processing efficiency.
174
175 As a real-world application, we used DisCo to explore the developmental heterogeneity of
176 intestinal organoids[31]. These polarized epithelial tissues are generated by intestinal stem cells in
177 3D matrices through a stochastic self-organization process, and mimic key geometric,
178 architectural and cellular hallmarks of the adult intestinal mucosa (e.g. a striking crypt-villus-like
179 axis)[31]. When grown from single stem cells, organoids of very different morphologies form under
180 seemingly identical *in vitro* conditions (**Figure 2A,** overview image in **Supplementary Figure 2A**).
181 Pooled tissue scRNA-seq data has shed light on the *in vivo*-like cell type composition of these
182 organoids[18–20,32], but cannot resolve inter-organoid heterogeneity. Critical for organoid

development is an early symmetry breaking event at Day 2 (16-32 cell stage) that is triggered by cell-to-cell variability and results in the generation of the first Paneth cell that is responsible for crypt formation[18]. Here, we were particularly interested in examining the emergence of heterogeneity between individual organoids subsequent to the symmetry breaking timepoint. To do so, we isolated single LGR5^{+} cells by FACS, and maintained them in a stem cell state using CHIR99021 and valproic acid (CV)[33]. On Day 3 of culture, CV was removed to induce differentiation. In total, we sampled 31 single intestinal organoids across four timepoints (Day 3 - 6) (**Figure 2A**). These organoids were selected based on differences in morphology (e.g. size variation, and cystic versus non-cystic morphologies), and may thus not constitute an unbiased sample of the population. Since Day 3 represents both differentiation Day 0 and the first sampling time point, we re-annotated the data accordingly (S0 – S3 replacing Day 3 – Day 6). During the co-encapsulation run, the number of encapsulated cells was noted and correlated to the number of barcodes retrieved, which was in approximate accordance (**Supplementary Figure 2B;** for an overview of the number of sequenced cells per organoid, see **Supplementary Table 1**). The even distribution of the number of reads mapping to ribosomal protein-coding genes and the observed low expression of heat shock protein-coding genes indicates that most cells were not affected by dissociation and on-chip processing (**Supplementary Figure 2C**).

To retrieve a first overview of overall cellular heterogeneity, we jointly visualized all 945 cells passing the quality thresholds through Uniform Manifold Approximation and Projection (UMAP). We found that our data was consistent with previously published pooled organoid scRNA-seq read-outs[19,32] since it revealed expected cell types including *Fabp1*-expressing enterocytes, *Muc2*-expressing goblet cells, *Reg3b*-positive Paneth cells, and *Olfm4*-expressing stem cells (**Figure 2B** and **2C**). In addition, a rare subset of cells, likely too few to form clusters, showed *ChgA* and *ChgB* expression, indicating the expected presence of enteroendocrine cells (**Supplementary Figure 2D**). Noteworthy, we found that batch effects are correctable since no

batch-based clustering was observed after correction (**Supplementary Figure 2E**). We also did not detect any clustering driven by cell quality, e.g. detected transcripts or mitochondrial transcripts (**Supplementary Figure 2C**). To further validate that batch effects between individual organoids can be corrected, we generated an independent dataset of an additional nine individual organoids (**Supplementary Figure 2F**). One of these nine organoids was split into two independent samples. Both of these two samples were processed with a 60-minute time delay in between. We found that the two halves of the split organoid were overlapping in the denominator UMAP (**Supplementary Figure 2G**), indicating that batch effects between individual organoids are indeed correctable. This includes potential batch effects that may be introduced by extended storage times. These findings support the cell type-resolving power of our DisCo platform (**Figure 2C,** extensive heatmap in **Supplementary Figure 2H,** and list in **Supplementary Table 2**).

In addition to the expected cell types, we observed a distinct cluster marked by high expression of Stem cell antigen 1 (*Sca1* or *Ly6a*). In depth analysis of marker genes showed high expression of *Anxa1* and *Clu* in the same cluster (**Supplementary Figure 2D**), and increased YAP-1 target gene expression (**Supplementary Figure 2I**), suggesting that these cells are most likely regenerative fetal-like stem cells[26,27,34]. Since the two remaining clusters did not show a striking marker gene signature, we resolved their identity by imposing temporal information on the data. This revealed that these clusters likely represent stem- and previously termed potentially intermediate cells (PIC)[35], given their occurrence at early developmental time points (**Figure 2D**). As expected, mature cell types were mostly present at later time points. To further leverage the temporal component in the DisCo data, we used slingshot trajectory analysis[36] to infer lineage relationships between cell types and to identify genes that may be of particular significance for waypoints along differentiation (**Figure 2E**). Beyond the previously utilized marker genes for cell type annotation, for example *Reg3b* and *Reg3g* for Paneth cells, additional markers that were validated in previous studies[37] were identified, such as *Agr2 and Spink4*, and *Fcgbp* for goblet

cells (**Figure 2F**). Overall, this suggests that the meta-data produced with our DisCo platform aligns with and expands prior knowledge.

Intriguingly, we observed the maintained presence of the *Ly6a*+ stem-cell population at S0, S1, and S3. Since cells with similar expression signatures were previously described under alternate culture conditions as belonging to a distinct organoid subtype termed spheroids[25], we next aimed to verify the presence of such spheroids among our sampled organoids and study their temporal behavior. To do so, we stratified our cells according to the individual organoids from which they were derived by mapping this information onto the reference scaffold (**Figure 3A**). Globally, this analysis revealed that the maturation seems to follow the expected pattern with early organoids (S0) mainly containing stem and Paneth cells, and older organoids (S1 – S3) differentiated cells like goblet cells and enterocytes. However, within single organoids, we found strong heterogeneity, revealing that *Ly6a*$^{+}$ cells were indeed present in a distinct subset of organoids, predominantly composed of these cells (S1a, S3e). Furthermore, images obtained prior to dissociation showed that *Ly6a*$^{+}$ cell-containing organoids (S3e) exhibited a larger, cystic like structure (**Supplementary Figure 3A**). To confirm the presence of *Ly6a*$^{+}$ organoids in our cultures, we utilized RNAscope (**Figure 3B,** controls **Supplementary Figure 3B**) to localize *Ly6a*, *Muc2,* and *Fabp1* expression in organoid sections. These analyses revealed canonical budding organoids, containing few *Muc2*$^{+}$ goblet cells and *Fabp1*$^{+}$ enterocytes, and *Ly6a*-expressing cells in spherical organoids that did not contain differentiated cell types such as enterocytes or goblet cells.

The presence of *Ly6a*$^{+}$ cells during the first day of sampling suggested that these cells constitute a second, *Lgr5*-independent stem cell population in the organoid culture. Using flow cytometry, we found that the majority of cells are either LGR5^{+} LY6A^{-} (24.5 %) or LGR5^{-} LY6A^{+} (3.3 %) with only a minority (0.4%) being double positive (**Figure 3C**). This finding, in combination with our

trajectory analysis (**Figure 2E** and **2F**), suggested that *Ly6a*$^{+}$ cells are capable of differentiating into organoids. To test this, we sorted and differentiated LGR5^{-} LY6A^{+} cells, revealing that both LGR5^{+} LY6A^{-} and LGR5^{-} LY6A^{+} cells give rise to organoids of similar morphological heterogeneity (**Figure 3D**). These results indicate that LGR5^{-} LY6A^{+} cells have full stem cell potential, comparable to that of previously described fetal-like stem cells[25]. Furthermore, the fact that LGR5^{-} LY6A^{+} cells did not display a propensity towards spheroid formation suggests that environmental conditions, e.g. variation in matrix stiffness, rather than the initial cell state dictate the formation of spheroids.

Besides the *Ly6a*$^{+}$ cell-enriched organoids, our data suggested the presence of additional organoid subtypes in the per organoid mappings (**Figure 3A**). The two most striking additional subtypes were three organoids that contained mostly enterocytes (S2c, S3a, S3d), and two that consisted predominantly of immature and mature goblet cells (S1b and especially S2f). The identity of the observed subtypes was further substantiated when visualizing the cell type abundance per organoid (**Figure 3E**), and marker gene expression in individual organoids (**Supplementary Figure 3C**). Similar to the spheroids, both subtypes showed aberrant morphologies, tending to be small and round, as compared to canonical organoids bearing a crypt-villus axis (e.g. S3c, **Supplementary Figure 3A**). To detect more subtle molecular differences, we used psupertime[38] to identify genes that are dynamically expressed during the development of individual organoids. This analysis revealed additional genes that are expressed in subsets of organoids, such as Gastric inhibitory polypeptide (*Gip*), Zymogen granule protein 16 (*Zg16*), Vanin 1 (*Vnn1*), and Defensin alpha 24 (*Defa24*) (**Supplementary Figure 3D**).

While organoids dominated by enterocytes were previously described as enterocysts[18], organoids displaying goblet cell hyperplasia, here termed “gobloids”, were so far to our knowledge unknown. To validate the existence of the uncovered organoid subtypes, we utilized RNAscope to localize

the expression of enterocyte (*Fabp1*) and goblet cell (*Muc2*) markers (**Figure 3F,** controls in **Supplementary Figure 3B**). In agreement with our data and prior research, we detected organoids that exclusively contained *Fabp1*$^{+}$ cells, most likely representing enterocysts. Most importantly, we were able to identify organoids that contained a high number of *Muc2*$^{+}$ goblet cells, confirming the existence of "gobloids".

Finally, to complement the intestinal organoid data and to provide a proof-of-principle for Disco's capacity to also *process in vivo*-derived small samples, we set out to analyze individual crypts that were isolated from the small intestine of adult C57BL/6J mice. However, we found that the dissociation of these crypts into single cells was more challenging than that of *in vitro* grown organoids. Utilizing the most efficient dissociation conditions among various tested ones, achieving efficiencies of up to 20% at elevated multiplet rates (**Supplementary Table 3** and **Material and Methods**), we analyzed 21 individual crypts involving 372 cells at a comparable cell recovery efficiency as for organoids (**Supplementary Figure 4A** and **Supplementary Table 1** for the number of sequenced cells). Next to individual crypts, we also utilized DisCo to generate a reference map of 775 cells derived from pooled crypts (bulk), which we integrated with the individual crypt cells to resolve their composition. This allowed us to identify distinct groups of cells: clusters marked by the expression of the cell cycle genes *Orc6* and *Top2a,* suggesting that these represent transit amplifying cells in G1/S- and G2/M-phase, respectively; two enterocyte clusters marked by *Fabp1* and *Apoa1* expression, and a goblet cell cluster marked by *Muc2* expression (**Figure 3G** and **Supplementary Figure 4B**). Most of these cell types were observed in bulk as well as individual crypt samples, except for enterocytes which were mainly detected in the bulk proportion. This likely reflects that bulk samples, in contrast to individual crypts, were not picked and scrutinized individually and thus possibly still contained residual villi (**Figure 3H**). Globally, the data overlapped with previously reported single cell data from bulk crypts[32], except for the lack of rare enteroendocrine cells and tuft cells with an expected abundance of only 1% in

bulk crypt isolates[32], and Paneth cells. However, we were able to identify the latter independent from clustering, namely by their gene expression signature (**Supplementary Figure 4C**).

Next to the expected cell types, we observed an additional cluster marked by the expression of *Clu* and *Anxa1*, which are established markers of regenerative, or revival stem cells.[26] Interestingly, we found three crypts that only contained these regenerative stem cells and that, providing an accurate compositional representation after dissociation, are thus depleted of other intestinal cell types (**Figure 3I,** all crypts in **Supplementary Figure 4D**). Since this observation aligned with our intestinal organoid (spheroid)-based findings, we next aimed to specifically explore whether spheroids and crypts contain comparable regenerative stem cells. To do so, we integrated the crypt data with the previously generated organoid data, yielding a common dataset of 2244 cells (**Figure 3J**, **Supplementary Figure 5A-C**). Strikingly, the regenerative stem cell cluster overlapped in the combined dataset, suggesting that this cell state can be recovered in both intestinal crypts and organoids, and that thus spheroids and regenerating crypts are compositionally comparable (**Figure 3K,** all crypts and organoids in **Supplementary Figure 5D**). Although caution is warranted when interpreting these results given the encountered dissociation issues, our findings indicate that some organoid heterogeneity recapitulates *in vivo* tissue heterogeneity, but also that crypts that predominantly contain regenerative stem cells are present in the homeostatic intestine. Altogether, our crypt data clearly support DisCo's capacity to profile *in vivo*-derived small, individual tissues, rendering the overall dissociation efficiency and no longer the processing efficiency the overall limiting factor.

Discussion

A key feature of our new DisCo approach is the ability to deterministically control the cell capture process. Despite lowering the throughput compared to stochastic droplet systems[2,3], our approach provides the advantage of being able to process low cell input samples at high efficiency and at a strongly decreased per cell cost (**Table 1**). Thus, we believe that the DisCo approach is filling an important gap in the scRNA-seq toolbox. Moreover, full control over the encapsulation process allows for continuous operation of our platform, which is offsetting to some extent the decreased throughput. Another critical feature of DisCo is the use of machine-vision to obtain full control of the entire co-encapsulation process including particle detection, particle positioning, particle droplet injection, and droplet volume. This enables the correct assembly of most droplets, virtually eradicating confounding factors that arise due to failed co-encapsulations[39,40]. In concept, DisCo is thus fundamentally different to passive particle pairing approaches such as traps[41–43] and, compared to these technologies, offers the advantage of requiring vastly simpler and reusable chips without suffering from cell/particle size and shape selection biases[15,44]. This renders the DisCo approach universally applicable to any particle co-encapsulation application[45,46], i.e. cell-cell encapsulations, with the only limiting factor being particle visibility. Providing further development, we envision that machine learning-based deterministic cell handling will ultimately enable targeted cell selection, e.g. by fluorescence or morphology, transforming DisCo into an end-to-end cell processor for samples with low-to-medium input samples.

To demonstrate DisCo's capacity to process small tissues/systems that were so far difficult to access experimentally, we have analyzed the cell heterogeneity of chemosensory organs from *D. melanogaster* larvae[47] and, as shown here, single intestinal organoids and crypts. It is thereby worth noting that, based on our handling of distinct tissues, we found that not DisCo itself, but rather cell dissociation has become the efficiency-limiting factor, a well-recognized challenge in

the field[48,49]. Indeed, substantial cell loss was a regular occurrence, even with optimized dissociation and processing strategies (see **Methods**).

scRNA-seq of individual organoids led us to uncover organoid subtypes of aberrant cell type distribution that were previously not resolved with pooled organoid scRNA-seq[18,19,32]. One subtype contained predominantly cells that were strikingly similar to previously described fetal-like stem cells or revival stem cells that occur during intestinal regeneration[26,27,34]. This subtype, previously described under alternate culture conditions as spheroid-type organoids[20,24,25], was identified here under standard organoid differentiation conditions, indicating that these organoids are capable of maintaining their unique state. We isolated LY6A-expressing cells and found that they readily give rise to canonical organoids, indicating that these cells are capable of providing a pool of multipotent stem-cells. Interestingly, in our proof-of-principle single intestinal crypt DisCo dataset, we identified crypts that largely consisted of cells with a similar regenerative gene expression signature. While crypts with these properties have been previously described upon injury, e.g. by irradiation,[26] our data suggests that such regenerating crypts are also present in the homeostatic intestine.

Of particular interest among the identified organoid subtypes was one that we termed “gobloid" given that it predominantly consists of immature and mature goblet cells. Since low Notch signaling is pivotal for the commitment of crypt base columnar (CBC) cells towards secreting progenitors, lack of Notch ligand-providing Paneth cells[50], may drive gobloid development[51]. However, failure to produce Paneth cells has previously been suggested as a mechanism underlying enterocyst development[18], which in principle requires high Notch signaling. Hence, we believe that our findings establish an important foundation to support further research on the emergence of gobloids and enterocysts from the still elusive PIC cells, providing an exciting opportunity to delineate lineage commitment factors of CBC cell differentiation.

In sum, we demonstrate that our DisCo analysis of individual intestinal organoids and crypts is a powerful approach to explore *in vitro* and *in vivo* tissue heterogeneity, and to yield new insights into how this heterogeneity arises. In comparison to established approaches such as automated microscopy[18,20], DisCo is magnitudes lower in experimental scale. Nevertheless, our intestinal scRNA-seq data enabled us to recapitulate previous findings, benchmarking DisCo, and most importantly, to uncover novel subtype entities, leveraging the key advantage of scRNA-seq, i.e. independence from *a priori* knowledge. Next to catalyzing research on other tissues or systems of interest, we believe that the technology and findings of this study will contribute to future research on intestinal organoid development and thus aid the engineering of more robust organoid systems. Furthermore, we believe that the utility of our presented approach extends to research on all developing multicellular organisms, and coupled with lineage tracing[52], will offer an entirely new perspective on interindividual variation. Finally, we expect this approach to be applicable to rare, small clinical samples to gain detailed insights into disease-related cellular heterogeneity and dynamics.

Acknowledgements

We thank Wanze Chen and Petra C. Schwalie for constructive discussions. We thank Virginie Braman for help in establishing intestinal organoid culture in our group, and Giovanni Sorrentino from Kristina Schoonjans' lab for valuable advice and support during organoid culture establishment. We also thank Luc Aeberli and Georges Muller from SEED Biosciences for cell sorting support. We thank the EPFL CMi, GECF, BIOP, FCCF, Histology core facility, SCITAS, and UNIL VITAL-IT for device fabrication, sequencing, imaging, sorting, histology, and computational support respectively. We particularly thank Jessica Sordet-Dessimoz for her support with the RNA-scope assay. This research was supported by the Swiss National Science Foundation Grant (IZLIZ3_156815) and a Precision Health & related Technologies (PHRT-502) grant to B. D., the Swiss National Science Foundation SPARK initiative (CRSK-3_190627) and

the EuroTech PostDoc programme co-funded by the European Commission under its framework programme Horizon 2020 (754462) to J. P., as well as by the EPFL SV Interdisciplinary PhD Funding Program to B. D. and E. A.. Y.S. is an ISAC Marylou Ingram scholar.

Contributions

BD, JB, and MB designed the study. BD, JB, MB, and JP wrote the manuscript. JB and RD designed and fabricated microfluidic chips. JB developed the machine-vision integration for DisCo. JB and MB benchmarked the system and performed all single-cell RNA-seq experiments. JP, JB, MB, WS, VG, and RG performed data analysis related to single organoid and crypt scRNA-seq experiments. JB, SR and MB performed all organoid and cell culture assays. JB, AC, JR, and RS, performed all imaging assays. MB and JR isolated intestinal crypts, MB picked and dissociated crypts. EA provided critical comments regarding microfluidic chip design and fabrication. MC provided critical comments on intestinal organoid scRNA-seq data analysis. ML provided critical comments regarding intestinal organoid scRNA-seq data and design of critical confirmation experiments. All authors read, discussed, and approved the final manuscript.

References

1. Tang, F. *et al.* mRNA-Seq whole-transcriptome analysis of a single cell. *Nat. Methods* **6**, 377–382 (2009).
2. Macosko, E. Z. *et al.* Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161**, 1202–1214 (2015).
3. Klein, A. M. *et al.* Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* **161**, 1187–1201 (2015).
4. Gierahn, T. M. *et al.* RNA sequencing of single cells at high throughput. *Nat. Methods* **14**, 395–398 (2017).
5. Han, X. *et al.* Mapping the Mouse Cell Atlas by Microwell-Seq. *Cell* **172**, 1091–1107 (2018).
6. Rosenberg, A. B. *et al.* Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding. *Science* **360**, 176–182 (2018).
7. The Tabula Muris Consortium. Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. *Nature* **562**, 367–372 (2018).
8. Han, X. *et al. Construction of a human cell landscape at single-cell level*. *Nature* (Springer US, 2020). doi:10.1038/s41586-020-2157-4
9. Saikia, M. *et al.* Simultaneous multiplexed amplicon sequencing and transcriptome profiling in single cells. *Nat. Methods* **16**, 59–62 (2019).
10. Stoeckius, M. *et al.* Cell Hashing with barcoded antibodies enables multiplexing and doublet detection for single cell genomics. *Genome Biol.* **19**, 1–12 (2018).
11. McGinnis, C. S. *et al.* MULTI-seq: sample multiplexing for single-cell RNA sequencing using lipid-tagged indices. *Nat. Methods* **16**, 619–626 (2019).
12. Gehring, J., Hwee Park, J., Chen, S., Thomson, M. & Pachter, L. Highly multiplexed single-cell RNA-seq by DNA oligonucleotide tagging of cellular proteins. *Nat. Biotechnol.* **38**, 35–38 (2020).
13. Hwang, B., Lee, J. H. & Bang, D. Single-cell RNA sequencing technologies and bioinformatics pipelines. *Exp. Mol. Med.* **50**, 1–14 (2018).
14. 10X Genomics. Chromium Single Cell 3' Reagent Kits v3 User Guide CG000183 Rev C. 1–55 (2018).
15. DeLaughter, D. M. The Use of the Fluidigm C1 for RNA Expression Analyses of Single Cells. *Curr. Protoc. Mol. Biol.* **122**, 1–17 (2018).
16. Wagner, D. E. *et al.* Single-cell mapping of gene expression landscapes and lineage in the zebrafish embryo. *Science* **360**, 981–987 (2018).
17. Packer, J. S. *et al.* A lineage-resolved molecular atlas of C. Elegans embryogenesis at single-cell resolution. *Science* **365**, (2019).
18. Serra, D. *et al.* Self-organization and symmetry breaking in intestinal organoid development. *Nature* **562**, 66–72 (2019).
19. Grün, D. *et al.* Single-cell messenger RNA sequencing reveals rare intestinal cell types. *Nature* **525**, 251–255 (2015).
20. Lukonin, I. *et al.* Phenotypic landscape of intestinal organoid regeneration. *Nature* **586**, 275–280 (2020).
21. Rossi, G., Manfrin, A. & Lutolf, M. P. Progress and potential in organoid research. *Nat. Rev. Genet.* **19**, 671–687 (2018).
22. Gjorevski, N. *et al.* Designer matrices for intestinal stem cell and organoid culture. *Nature* **539**, 560–564 (2016).
23. Brassard, J. A. & Lutolf, M. P. Engineering Stem Cell Self-organization to Build Better Organoids. *Stem Cell* **24**, 860–876 (2019).
24. Mustata, R. C. *et al.* Identification of Lgr5-Independent Spheroid-Generating Progenitors of the Mouse Fetal Intestinal Epithelium. *Cell Rep.* **5**, 421–432 (2013).
25. Yui, S. *et al.* YAP / TAZ-Dependent Reprogramming of Colonic Epithelium Links ECM Remodeling to Tissue Regeneration. *Cell Stem Cell* **22**, 35–49 (2018).
26. Ayyaz, A. *et al.* Single-cell transcriptomes of the regenerating intestine reveal a revival stem cell. *Nature* **569**, 121–125 (2019).
27. Roulis, M. *et al.* Paracrine orchestration of intestinal tumorigenesis by a mesenchymal niche. *Nature* **580**, 524–529 (2020).
28. Unger, M. A., Chou, H. P., Thorsen, T., Scherer, A. & Quake, S. R. Monolithic microfabricated valves and pumps by multilayer soft lithography. *Science* **288**, 113–116 (2000).

29. Biocanin, M., Bues, J., Dainese, R., Amstad, E. & Deplancke, B. Simplified Drop-seq workflow with minimized bead loss using a bead capture and processing microfluidic chip. *LoC* **19**, 1610–1620 (2019).
30. Zhang, X. *et al.* Comparative Analysis of Droplet-Based Ultra-High-Throughput Single-Cell RNA-Seq Systems. *Mol. Cell* 1–13 (2018). doi:10.1016/j.molcel.2018.10.020
31. Sato, T. *et al.* Single Lgr5 stem cells build crypt – villus structures in vitro without a mesenchymal niche. *Nature* **459**, 262–266 (2009).
32. Haber, A. L. *et al.* A single-cell survey of the small intestinal epithelium. *Nature* **551**, 333–339 (2017).
33. Yin, X. *et al.* Niche-independent high-purity cultures of Lgr5 + intestinal stem cells and their progeny. *Nat. Methods* **11**, 106–112 (2014).
34. Gregorieff, A., Liu, Y., Inanlou, M. R., Khomchuk, Y. & Wrana, J. L. Yap-dependent reprogramming of Lgr5+ stem cells drives intestinal regeneration and cancer. *Nature* **526**, 715–718 (2015).
35. Battich, N. *et al.* Sequencing metabolically labeled transcripts in single cells reveals mRNA turnover strategies. *Science* **367**, 1151–1156 (2020).
36. Street, K. *et al.* Slingshot : cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics* **19**, 1–16 (2018).
37. Birchenough, G., Johansson, M., Gustafsson, J., Bergstrom, J. & Hansson, G. C. New developments in goblet cell mucus secretion and function. *Mucosal Immunol.* **8**, 712–719 (2015).
38. Macnair, W. & Claassen, M. psupertime: supervised pseudotime inference for single cell RNA-seq data with sequential labels. *bioRxiv* 622001 (2019). doi:10.1101/622001
39. Lareau, C. A., Ma, S., Duarte, F. M. & Buenrostro, J. D. Inference and effects of barcode multiplets in droplet-based single-cell assays. *Nat. Commun.* **11**, 1–9 (2020).
40. Lun, A. T. L. *et al.* EmptyDrops: Distinguishing cells from empty droplets in droplet-based single-cell RNA sequencing data. *Genome Biol.* **20**, 1–9 (2019).
41. Chung, M., Núñez, D., Cai, D. & Kurabayashi, K. Deterministic droplet-based co-encapsulation and pairing of microparticles via active sorting and downstream merging. *Lab Chip* **17**, 3664–3671 (2017).
42. Cheng, Y. H. *et al.* Hydro-Seq enables contamination-free high-throughput single-cell RNA-sequencing for circulating tumor cells. *Nat. Commun.* **10**, 1–11 (2019).
43. Zhang, M. *et al.* Highly parallel and efficient single cell mRNA sequencing with paired picoliter chambers. *Nat. Commun.* **11**, 1–13 (2020).
44. Pollen, A. A. *et al.* Low-coverage single-cell mRNA sequencing reveals cellular heterogeneity and activated signaling pathways in developing cerebral cortex. *Nat. Biotechnol.* **32**, 1053–1058 (2014).
45. Dura, B. *et al.* Profiling lymphocyte interactions at the single-cell level by microfluidic cell pairing. *Nat. Commun.* **6**, 1–13 (2015).
46. Gérard, A. *et al.* High-throughput single-cell activity-based screening and sequencing of antibodies using droplet microfluidics. *Nat. Biotechnol.* **38**, 715–721 (2020).
47. Maier, G. L. *et al.* Multimodal and multisensory coding in the Drosophila larval peripheral gustatory center. *bioRxiv* (2020). doi:10.1101/2020.05.21.109959
48. Haque, A., Engel, J., Teichmann, S. A. & Lönnberg, T. A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications. *Genome Med.* **9**, 1–12 (2017).
49. Denisenko, E. *et al.* Systematic assessment of tissue dissociation and storage biases in single-cell and single-nucleus RNA-seq workflows. *Genome Biol.* **21**, 1–25 (2020).
50. Sato, T. *et al.* Paneth cells constitute the niche for Lgr5 stem cells in intestinal crypts. *Nature* **469**, 415–418 (2011).
51. Van Es, J. H. *et al.* Notch/γ-secretase inhibition turns proliferative cells in intestinal crypts and adenomas into goblet cells. *Nature* **435**, 959–963 (2005).
52. McKenna, A. *et al.* Whole-organism lineage tracing by combinatorial and cumulative genome editing. *Science* **353**, (2016).

Table 1

(**Left Subtable**) Performance summary of established scRNA-seq platform technologies. Performance metrics were derived from the literature. Noteworthy, as for lack of consensus experiments, metrics represent different values. Furthermore, the cost per cell is calculated for 100 cells (output: 100 single cells that are successfully processed, thus not incorporating platform-specific processing inefficiencies; input: a sample of 100 total cells that are processed on the respective system, hence considering platform-specific processing inefficiencies) to match the sample size utilized for DisCo experiments. (References and calculation of metrics are detailed in the Material and Methods section). (**Right Subtable**) Performance metrics calculated for the DisCo system as presented in this study.

Approach	**Droplets (stochastic)**			**FACS & plate based**		**Traps**	**Microwells**			**Droplets (deterministic)**
Technology	10X Chromium	inDrop	Drop-seq	Smart-seq2	Cel-seq2	Fluidigm C1	iCell 8	Seq-well		DisCo (this study)
Min input	500 (HT)/100 (LT)	1,000	50,000	10,000	10,000	<50	1,600	400		< 50
Efficiency	45%*/30%*	25%*	2.3%*	-	-	30 - 45%*	43%**	30%*		75%*
$/cell (100 output cells)	$20/$5.9	$2.1	$6	$10.6	$3.6	$29 (96 cells)	$5	$2.2		$1
$/cell (100 input cells)	$44.4/$19.8	$8.4	$260.9	-	-	$62.2	$11.6	$7.5		$1.3
Additional remarks or limitations	Multiplexing possible on HT, yet requires multiple washing procedures[10,11], and thus substantial efficiency losses expected.			Fluorescent labeling necessary Expensive to scale up (automation)		Size-selective properties[15, 44]	High initial acquisition cost			

Efficiency estimates: * **including** cell capture efficiency; ** **excluding** cell capture efficiency

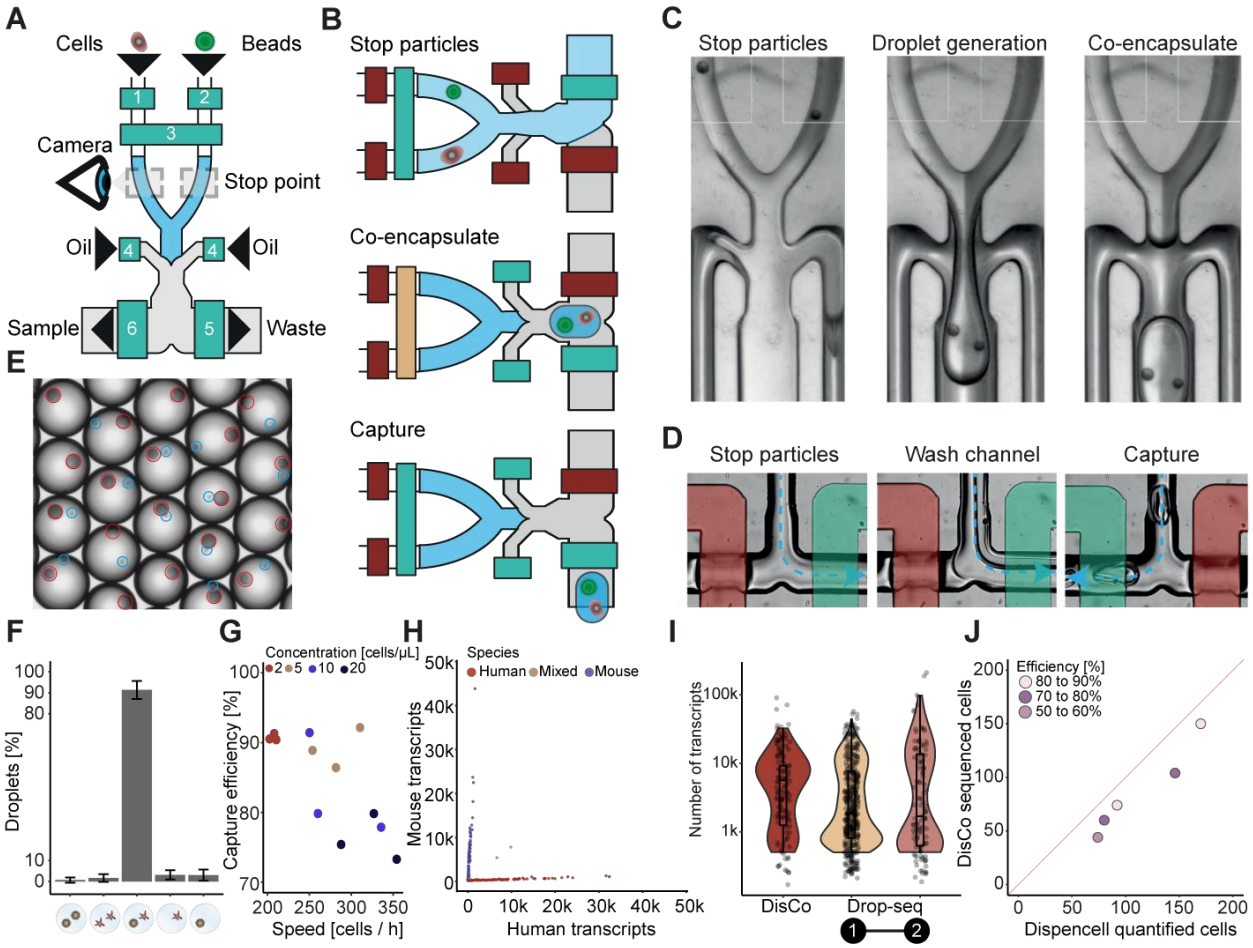


Figure 1. Overview and critical feature assessment of the deterministic co-encapsulation (DisCo) system: (**A**) Schematics of the DisCo microfluidic device. The device contains three inlet channels for cells, beads, and oil, and two outlets for waste and sample liquids. All inlets and outlets are augmented with Quake-style microvalves (green boxes): 1. cell valve, 2. bead valve, 3. dropleting valve, 4. oil valve, 5. waste valve, 6. sample valve. The device is continuously monitored by a high-speed microscopy camera to detect and coordinate the placement of particles at the Stop point. (**B**) Illustration of the particle co-encapsulation process on the DisCo device. Initially, two particles (here a bead and a cell) are stopped (Stop particles) in close proximity to the channel junctions by closing the channel valves (red: closed, green: open). Next, by pressurizing the dropleting valve (yellow), both particles are ejected into the junction point, and the droplet is sheared by opening the oil valve (Co-encapsulate). Finally, the produced droplet is captured in the Sample channel (Capture). (**C**) The co-encapsulation process of two beads and droplet generation as observed on chip. Dyed liquids were used to examine the liquid interface of the carrier liquids. Channel sections with white squares are 100 µm wide. (**D**) The droplet capture process as observed on-chip. Valves are highlighted according to their actuation state (red: closed, green: open). While particles are stopped, excess buffers are discarded through the waste channel and the channel is flushed with oil prior to droplet capture. Upon co-encapsulation, the waste valve is closed, the sample valve opened, and the produced droplet captured in the Sample channel. (**E**) Images of DisCo droplet contents. Cells (blue circle) and beads (red circle) were co-encapsulated, and captured droplets imaged. Mean bead-size is approximately 30 µm. (**F**) Droplet occupancy of DisCo-processed cells and beads for cell concentrations ranging from 2 to 20 cells per µl (total encapsulations n = 1203). Error bars represent the standard deviation. (**G**) Cell capture efficiency and speed for varying cell concentrations (total encapsulations n = 1203). Cells were co-encapsulated with beads at concentrations ranging from 2 - 20 cells per µl, and co-encapsulation events quantified by analyzing recordings of the process. (**H**) DisCo scRNA-seq species separation experiment. HEK 293T and murine pre-adipocyte iBA cells were processed with the DisCo workflow for scRNA-seq, barcodes merged, and species separation visualized as a Barnyard plot. (**I**) Comparison of

detected UMIs per cell of conventional Drop-seq experiments. UMIs per cell from HEK 293T data for conventional Drop-seq experiments ([1] - from Biočanin, Bues *et al.* 2019[29] and [2] - from Macosko *et al.* 2015[2]), compared to the barcode-merged HEK 293T DisCo data. Drop-seq datasets were down-sampled to comparable sequencing depth. Box elements are described in the **Materials and Methods** section. (**J**) Total cell processing efficiency of DisCo at low cell inputs. Input cells (HEK 293T) ranging from 74 to 170 were quantified with the Dispencell system. Subsequently, all cells were processed with DisCo, sequenced, and quality filtered (> 500 UMIs). The red line represents 100% efficiency, and samples were colored according to the recovery efficiency after sequencing.

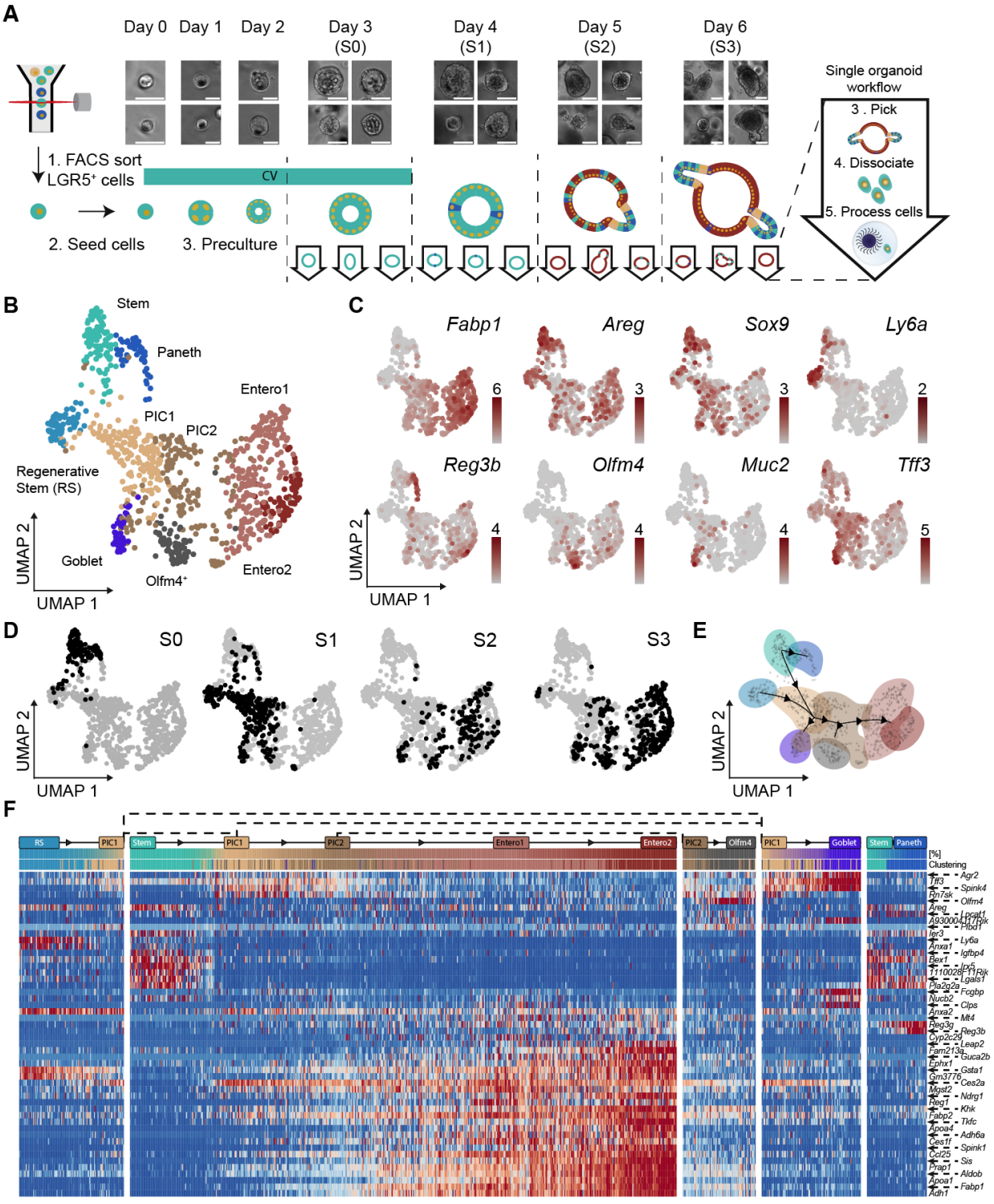


Figure 2. Utilizing DisCo to map intestinal organoid cell heterogeneity along development: (**A**) Overview of the experimental design for DisCo'ing individual organoids. Single LGR5⁺ intestinal stem cells were isolated via FACS and precultured for 3 days under stem cell maintenance conditions (ENR CV Day 0 to 3). On Day 3, CV was removed from the culture, and organoids differentiated under ENR conditions for up to 3 days. For each day during development (S0 - S3), individual organoids were isolated, dissociated, and processed on the DisCo platform. Representative bright-field imaging examples of

individual organoids for each day are shown on top. Scale bar: 25 µm (Day 0 - 2) 50 µm (Day 3 - 6). (**B**) UMAP embedding of all sequenced cells. All 945 processed cells from 31 organoids were clustered with k-means clustering, after which clusters were annotated according to specific marker gene expression. (**C**) UMAP-based visualization of the expression of specific markers that were used for cluster annotation. (**D**) Temporal occurrence of cells. Cells are highlighted on the UMAP embedding according to sampling time point (S0 - S3). (**E**) Developmental trajectory based on the cluster annotation and the sampling time point derived by slingshot[36]. Cells were annotated in accordance with clustering in (B). (**F**) Heat map of differentially expressed genes along the waypoints of the trajectory. Waypoints are annotated in accordance with cell clustering as in (B). Cluster abbreviations: Stem cells (Stem), Regenerative stem cells (RS), Potential intermediate cells (PIC)[35], Enterocytes cluster 1/2 (Entero1/2).

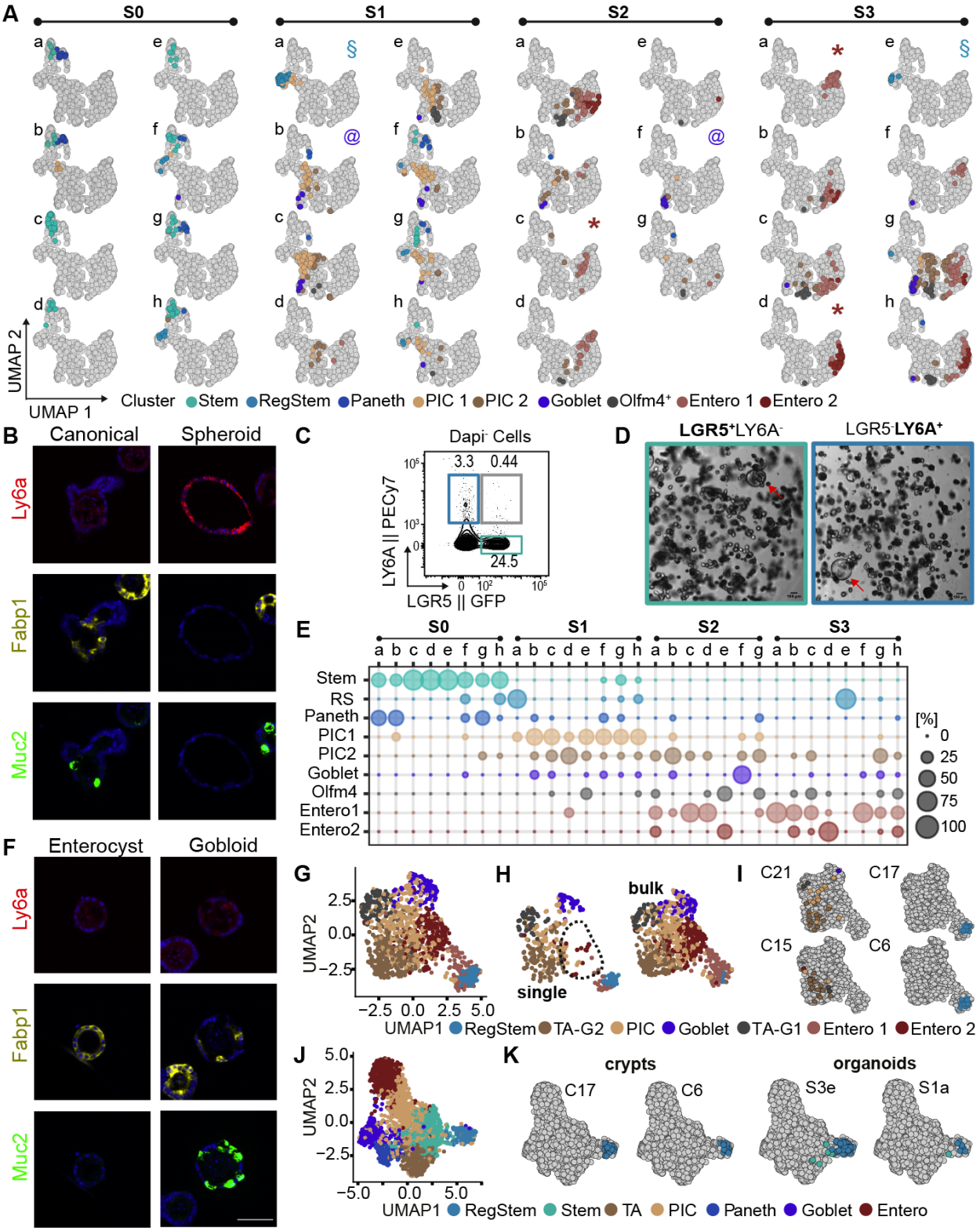


Figure 3. Cell type distribution and marker gene expression across individual intestinal organoids and crypts: (**A**) Projection of cell types onto 31 individual organoids. Cells per single organoid were colored according to their global clustering and highlighted on the UMAP embedding of all sequenced cells. Projections are grouped according to their sampling time. Manually classified organoids were annotated with the following symbols: "*" enterocysts, "§" spheroids, "@" gobloids. **(B)** *in situ* RNA detection of *Ly6a*,

Fabp1, and *Muc2* expression. A representative canonical and *Ly6a*-expressing organoid is displayed. Scale bar (displayed in F): 50 µm. (**C**) Surface LY6A and LGR5-GFP expression under ENR CV conditions. The dot plot depicts LGR5-GFP and LY6A expression in organoid-derived single cell suspensions. The numbers indicate frequencies (%). (**D**) Culturing outcomes of $LGR5^+$ cells and $LY6A^+$ cells. Single $LGR5^+$ $LY6A^-$ and $LGR5^-$ $LY6A^+$ cells were isolated by FACS and seeded in Matrigel. Cells were cultured as depicted in **Figure 2A** and imaged using bright-field microscopy at S3. Red arrows point to spheroid morphologies. Scale bar: 100 µm. (**E**) Dotplot depicting the distribution of annotated cell types per organoid. Dot size depicts the percentage of cells associated to each cluster per organoid. **(F)** *in situ* RNA detection of *Fabp1* and *Muc2* expression. Selected images resembling the enterocyst and gobloid subtypes. Scale bar: 50 µm. **(G)** UMAP embedding of all cells collected from bulk and individual crypts. All 775 processed cells from bulk and 372 cells from individual crypts were clustered with k-means clustering, after which clusters were annotated according to marker gene expression. **(H)** UMAP depiction of cells derived from bulk or individual crypts. The dotted line highlights the enterocyte cluster. **(I)** UMAP from G) superimposed with cells from exemplary single crypts. **(J)** UMAP embedding of all sequenced cells obtained from intestinal organoids (**Figure 2**) and crypts. All 2244 processed cells were clustered with k-means clustering, after which clusters were annotated according to marker gene expression. **(K)** UMAP from J) stratified by exemplary single crypts and organoids that are largely composed of regenerative stem cells. Enterocytes (Entero), PIC (Potential intermediary cells), RegStem, (Regenerative Stem), TA (Transit amplifying cells; G1: G1/S and G2: G2/M cell cycle phase).

Supplementary Materials:

1. Materials and Methods
2. Supplementary Figures 1-5
3. Supplementary Tables 1-3

1. Materials and Methods

System comparison metrics

Performance metrics for (Supplementary Table 1) were calculated the following ways:

- **Minimum cell input estimates:** The minimum cell input values were derived from the following sources: **10X Chromium HT/LT**[1]**:** Lowest cell input number from the 10X Chromium manual (HT: CG000183 Rev C, LT: CG000399 Rev B); **inDrop**[2]**:** Lowest numbers mentioned in the 1CellBio manual (Single Cell Encapsulation Protocol, Version 2.4); **Drop-seq**[3]**:** Lowest numbers utilized in Zhang et al. 2019[4]. It is likely that lower cell numbers can be processed, yet Drop-seq has been suggested to be used "When the sample is abundant" by Zhang et al. 2019[4]; **FACS-based methods**[5,6]**:** input limits as described by Hwang et al. 2018[7]; **Fluidigm C1**[8]**:** Lowest cell input number used for benchmarking experiments used in this work; **Wafergen iCell8**[9]**:** Lowest cell numbers were derived from the iCell manual (CELL8 Single-Cell ProtocolD07-000025 Rev. C). According to the manual, 80 µL of 0.02 cells/nL suspension are prepared for dispensing; **Seq-well**[10]**:** The lowest cell number used for capture in Gierahn et al. 2017[10]. **Disco:** The lowest cell number processed in this study.
- **Efficiency estimates:** Efficiency estimates were derived from varying sources and represent different efficiencies. The efficiencies for **10X Chromium HT**, **inDrop**, and **Drop-seq** were derived from Zhang et al. 2019[4] from quantified cellular inputs (> 1000 cells) and sequenced cells passing quality thresholds. Since these efficiencies stem from experiments that were performed with optimized cell inputs, we can assume lower efficiencies when processing low cell inputs (< 1000). For the **10X Chromium LT** kit, efficiencies were derived from the user manual CG000399 Rev B. The efficiency for the **Fluidigm C1** system was determined in this work (results shown in Supplementary Figure 1L). For the **Wafergen iCell8** system, an efficiency estimate was derived from Wang et al. 2019[11] and represents the conversion efficiency from captured to sequenced cells passing quality thresholds, thus it does not include cell capture inefficiencies. The efficiency for **Seq-well** was derived from Gierahn et al. 2017[10] at 400 cells input and represents an inferred efficiency from quantified cell input to sequenced cells passing quality thresholds. Specifically, the

library conversion efficiency, i.e. the percent of captured cells identified in the sequencing data passing quality thresholds, was calculated based on the species-mixing experiment involving 10,000 input cells. The library conversion efficiency, in combination with capture efficiencies at 400 cells, was utilized to determine the efficiency at low cell numbers. Hence, this is inferred from quantified cellular inputs to sequenced cells passing quality thresholds. **DisCo:** The efficiencies were derived in this study and represent mean efficiencies for low cell inputs (50 - 200), from quantified cell input to sequenced cells passing quality thresholds.

- **Cost per cell estimates:** Two cost estimate numbers are listed for 100 cells: i) the cost for 100 cells not considering system efficiencies ($/cell, 100 output cells), and ii) the cost for 100 input cells considering the listed efficiencies ($/cell, 100 input cells). Run costs for **Smart-seq2**, **Cel-seq2**, **inDrop**, **Drop-seq**, and **Seq-well** were derived from Ding et al. 2020 (Supplementary Table 8)[12]. Run costs for **10X Chromium HT**, **Fluidigm C1** (96), and **Wafergen iCell8,** were derived from Wang et al. 2019 (Table 2)[11]. Costs for **10X Chromium LT** were derived from the 10X price list of the EPFL sequencing core facility (GECF). For the **Wafergen iCell8** it was assumed that 8 samples (one per dispensing nozzle) can be processed on one chip in parallel, thus decreasing the costs by a factor of 8. The **DisCo** cost estimate includes reagents for library generation, i.e. the costs for beads, oil, reverse-transcription reaction, exonuclease treatments, PCR reaction, and library preparation (Nextera XT).

Physical setup

Chips were mounted on an IX51 inverted microscope (Olympus). Each chip was monitored with an XiC (Ximea, MC031MG-SY-UB) camera, interfaced with a computer with the following specifications: Windows 10 Enterprise (Microsoft) operating system, Ryzen Threadripper 1950X processor (AMD), 32 GB RAM memory. Solenoid valves were controlled via the NI USB-6501 controller (National Instruments). The output signals from the controller were amplified with a ULN2803 IC (Texas Instruments), and connected to solenoid valves (Festo, MHA1-M1H-3/2O-0,6-HC). An OB1 Mk3 pressure controller (Elveflow) was used for proportional pressure regulation.

Machine-vision software

The software for cell detection and coordination was implemented in C++. Camera images were obtained with the XiApi library (version 4.15). Images were processed in real-time using the OpenCV computer vision library (version 3.4). A schematic visualization of the particle detection algorithm is depicted in **Supplementary Figure 1B**. Briefly, a detection ROI was extracted by cropping after which a gaussian blur was applied to the resulting image. Two subsequent images were subtracted, and the resulting image converted to a binary image by intensity thresholding. The binary image was dilated to fill potential holes. Finally, contours were detected using the *findContours* function, and classified for area and circularity. Upon particle detection, the particles were properly positioned by valve oscillation and monitoring of the ROI at the target zone (**Supplementary Figure 1C**). Once two particles were positioned in their respective target zones, particles were co-ejected by pressurization of the dropleting valve, and the droplet was sheared by actuation of the oil valve.

Microfluidic chip design and fabrication

The design of the microfluidic chip for deterministic co-encapsulation is presented in **Supplementary Figure 1A**. Chips were designed using Tanner L-Edit CAD software (Mentor, v 2016.2). 5-inch chromium masks were exposed in a VPG200 laser writer (Heidelberg instruments) for both the control and flow layer. Masks were developed using an HMR 900 mask processor (Hamatech). For the control layer, a thick SU8 photoresist layer was deposited with an LSM-200 spin coater (Sawatec), exposed on a MJB4 single side mask aligner (SussMicroTec), and manually developed. The SU8 processing steps were carried out according to manufacturer's instructions for the 3010 series (Microchem). For the flow layer, wafers were produced using AZ40 XT (Microchem) positive photoresist on the ACS200 coating and developing system (Gen3, SUSS MicroTec). Developed master-wafers were reflowed for 45 - 75 seconds at 120^{o}C on a hotplate until channels appeared round under an inspection microscope. The control layer master-wafers were used as molds for PDMS chips after passivation with 1 % silane dissolved in HFE. For the flow layer, master-wafers were used to generate replica molds for chip production. To this end, the primary replica mold was obtained by mixing PDMS:Curing-Agent at 10:1 using a centrifugal mixer (Thinky), degassing for 15 minutes, and curing for 60 minutes at 80°C. The PDMS-based primary replica mold was then sylanized

and subsequently used to obtain secondary replica molds utilized for PDMS flow layer production. The PDMS flow layer was fabricated PDMS:Curing-Agent at 5:1, degassed and cured at 80°C for 30 minutes. The control layer was fabricated by spin coating PDMS:Curing-Agent at 20:1 on the flow layer waver at 650 rpm for 35 seconds with 15 seconds ramp time followed by baking at 80°C for 30 minutes. Cured PDMS was then cut from the flow layer secondary replica mold and flow layer inlet holes were punched with a 0.5 mm diameter biopsy punch. The two PDMS layers were manually aligned and bonded at 80°C for at least 60 minutes. Assembled and cured PDMS chips were cut from the molds and control layer inlet holes were punched. Finally, chips were oxygen plasma activated (45 seconds at ~500 mTorr O2) and bonded to a surface activated glass slide followed by incubation at 80°C for at least 2 hours. Materials and reagents are listed in the Material and reagent list, point 1.

Microfluidic device handling

Prior to use, the microfluidic chip was placed on an inverted microscope and control layer inlets were connected to solenoid valves with water primed tygon tubing. Control layer channels were primed with dH_2O in tygon tubing for ~10 minutes by pressurizing the solenoid valves. If the chip was being used for the first time, cell, bead, and dropleting on-chip valves were equilibrated by oscillation of the corresponding solenoid valves for at least 10 minutes at 2 actuations per second. After priming, the dropleting valve was connected to an OB1 (Elveflow) pressure regulator for proportional actuation. The flow layer was connected the following way: oil, bead, cell inlets and sample outlet to Prot/Elec gel loading tips; waste outlet to tygon tubing terminating in a falcon tube. For inlet pressurization of the Prot/Elec gel loading tip connected inputs, the bead and cell inlets were connected to the OB1 pressure regulator. The oil inlet was continuously pressurized at 1.7 psi. Cell, bead, and oil Prot/Elec tips were filled with cell buffer, bead solution, and oil, respectively. Subsequently, the chip was primed in the following order: 1. cell channel, 2. bead channel, 3. oil channel. After priming, the bead and cell channels were washed for 5 - 10 minutes by running the solutions at low pressure. All priming and washing solutions were directed in the waste outlet. Finally, the sample outlet was primed with oil. Stuffer droplets, containing lysis buffer and RNase inhibitor, were generated on a Drop-seq chip[13] and added on top of the oil-primed sample outlet tip without introducing air bubbles. Materials and reagents are listed in the Material and reagent list, point 2.

cDNA generation and library preparation

After bead-cell in droplet co-encapsulation, the gel loading tip containing the sample droplets were transferred to a bead collection chip inlet[13] (cp-chip). Droplets in the tip were flushed to a bead collection chip. Subsequent to bead capture, washing was performed as in the Drop-seq protocol with SSC and reverse transcription buffer directly on the cp-chip. Reverse transcription solution was added to the beads in the recovery chip, and the recovery chip was placed on a heating block to perform first strand cDNA synthesis (RT) for 90 minutes at 42^{o}C. After the RT reaction, beads were washed on the recovery chip with TE-SDS once, with TE-TW twice, and with Tris once. The beads were treated with Exonuclease I for 45 minutes at 37^{o}C to remove single-stranded oligonucleotides on the beads. After Exonuclease I treatment, beads were washed with TE-SDS once, with TE-TW twice (as after RT). Beads were then eluted from the recovery chip in dH_2O. cDNA was amplified for 18 – 23 cycles using Kapa HiFi Hot start ready mix. cDNA was purified with CleanPCR magnetic beads (0.6X ratio) to remove small cDNA fragments and primers. The cDNA concentration was measured using Qubit, and cDNA quality was assessed using a Fragment Analyzer (Agilent). cDNA was tagmented with in-house Tn5[14] for 6 minutes at 55^{o}C. Next, the reaction was stopped with SDS and the tagmented library was amplified for 15 cycles using Kapa HiFi kit. Libraries were then purified using CleanPCR magnetic beads (0.6X ratio) and quantified using Qubit HS kit and Fragment analyzer (Agilent). Finally, size-selected and purified libraries were sequenced on a NextSeq 500 system (Illumina) following recommendations from the original Drop-seq protocol (20 bp for read 1 and 50 bp for read2)[15]. Material and reagents are listed in the Material and reagent list, points 3 - 10.

Mammalian cell culture handling for the species mixing experiment

For benchmarking the DisCo platform, HEK 293T (ATCC Cat. No. SD-3515) and murine brown preadipocyte cells (iBA; provided by Prof. Christian Wolfrum's laboratory, ETH Zürich) were used. Cells were cultured to 90% confluency in Glutamax DMEM supplemented with FBS and penicillin-streptomycin. Prior to use, cells were washed with PBS, dissociated with Trypsin-EDTA, washed with cell wash buffer and counted with Trypan blue live-dead stain using a Countess cell counter (Invitrogen). Cells were mixed

in a 1:1 ratio, adjusted to 20 cells/µL, re-suspended in cell loading buffer, and finally loaded on the DisCo chip. Material and reagents are listed in the Material and reagent list, point 11.

Droplet content and co-encapsulation performance quantification

As for conventional DisCo runs, experiments were set up with Chemgen beads and varying concentrations of HEK 293T cells. Approximately 100 co-encapsulations were performed and recorded. The recorded video data was manually reviewed and droplet contents and passing cells and beads counted (**Figure 1F**).

Benchmarking DisCo efficiency using the DISPENCELL platform

To benchmark single-cell recovery efficiencies throughout the complete DisCo workflow, we quantified HEK 293T (ATCC Cat. No. SD-3515) cells utilizing the DISPENCELL pipetting robot (SEED Biosciences SA). Prior to use, HEK 293T cells were diluted to 20 cells/µL. Cells were loaded into the DISPENCELL tip and then dispensed directly into a Prot/Elec gel loading tip containing cell loading buffer. Cells were then processed with DisCo and libraries prepared as described above.

Benchmarking the Fluidigm C1

To benchmark the cell recovery efficiency of the Fluidigm C1 platform, HEK 293T (ATCC Cat. No. SD-3515) cells were diluted with Suspension Reagent (Fluidigm) to reach approximately 10, 20, 40 cells per µL. The obtained suspensions were generated separately from the same stock and then quantified in triplicate using microscopy by examining a volume of 2 x 2.5 µL of the suspension between two coverslips. Counts of all triplicates were averaged to determine the cell input for 5 µL of Cell Mix, and the same volume was subsequently loaded on the C1 IFC. The experiment on the C1 machine was performed according to the kit's manual: "SMART-Seq v4 Ultra Low Input RNA Kit for the Fluidigm C1 System, IFCs User Manual" (Clontech Laboratories, Inc.) using 10 – 17 µm 96-trap C1 IFC OpenApp chips. The protocol was run on SMART-Seq v4 (1861x/1862x/1863x) programs on the C1 machine. To verify successful loading and cell trapping, traps were examined using a Cell xCellence (Olympus) microscope. Final cDNA was quantified using the PicroGreen dsDNA assay and then tagmented using the Nextera XT library preparation kit according to the manufacturer's instructions.

Material and reagent list for benchmarking the Fluidigm C1: For single-cell chip loading and priming, the C1 Single-Cell mRNA Seq HT Reagent kit v2 (Fluidigm, SKU 101-3473) was used as well as 10-17 µm 96-trap C1 IFC (Fluidigm, SKU 100-8134). For cDNA generation, a SMART-Seq v4 Ultra Low Input RNA Kit (Clontech Laboratories Inc, 634888) was used. cDNA was quantified using the PicoGreen HS dsDNA assay (Invitrogen, P11496) and then libraries were generated using Nextera XT (Illumina, FC-131-1096) and the Nextera XT index kit set A (Illumina, FC-131-2001).

Temporal batch effect experiment

A single cell suspension of HEK 293T (ATCC Cat. No. SD-3515) was loaded on the system as described above, and the remaining volume stored on ice. After 20 minutes, generated droplets were evacuated from the system and sequencing libraries prepared. A new cell loading tip was inserted into the sample outlet port and the experimental run was resumed. The previous steps were repeated after 40 and 60 minutes. After 120 minutes, the system was loaded with cells stored on ice and cells captured for approximately 20 minutes. Subsequently, droplets were evacuated from the system, cDNA was generated, and sequencing libraries prepared as described above. The former steps were repeated for cells stored for 180 minutes on ice.

For the material and reagent list for the temporal batch effect experiment, we refer to the section "mammalian cell culture handling for the species mixing experiment".

Mouse intestinal organoid culture and handling

Isolation of the Lgr5-eGFP^{+} stem cells and initial culture was performed as previously described[16]. For the developmental time-course experiments, organoids were dissociated to single cells, live Lgr5^{+}-eGFP cells isolated using a FACS ARIA II (BD) and embedded in Matrigel. After Matrigel polymerization, cells were cultured in ENR CV medium supplemented with thiazovivin ROCK inhibitor. Growth factors (E, N, R, C, V) were replenished after 2 days of culture. At Day 3 of culture, a full medium change was performed to differentiation growth medium (ENR only). At Day 5, growth-factors (E, N, R) were replenished. Organoids were sampled at Day 3 (S0), prior to the medium change, at Day 4 (S1), at Day 5 (S2), and at Day 6 (S3).

Single organoids were collected by dissolving Matrigel with ice-cold Cell Recovery Solution for approximately 5 minutes, while carefully pipetting up and down with a 1000 µL pipette. Subsequently, single organoids were isolated by hand-picking after which they were transferred to a Nunc microwell culture plate with single organoid dissociation mix. Single organoids were dissociated by combining trituration using siliconized pipette tips every 5 minutes and incubation at 37°C for 15 minutes. Following dissociation, cell suspensions were diluted in cell loading buffer in the loading tip connected to the DisCo chip. Materials and reagents are listed in the Material and reagent list, points 12 - 16.

Intestinal organoids were cultured in Matrigel (Corning, 356230) with organoid base medium (described in point 13) supplemented with ENR (+ CV where indicated) and ROCK inhibitor (where indicated, Sigma, Y0503). Organoid base medium was prepared using DMEM/F12 (Gibco, 11320033), Hepes (100 mM, Gibco, 15630056), penicillin-streptomycin (100 U/mL, Gibco, 15140122), B27 supplement (1 µM, Gibco, 17504-044), N2 supplement (1 µM, Gibco, 17502001), and N-Acetyl-L-cysteine (1 µM, Sigma, A9165). ENR medium was prepared using base medium (as above), EGF (E, 50 ng/mL, LifeTechnologies, PMG8043), mNoggin (N, 100 ng/mL, produced in-house), R-spondin (R, 1 µg/mL, produced in-house). ENR CV medium was prepared with addition of CHIR (C, 3 µM, CalBiochem, CHIR99021), and Valproic acid (V, 3 mM, Sigma P4543) to the ENR medium. Single organoid, single cell dissociation mix was prepared using PBS (Gibco, 14190-094), *B. licheniformis* protease (10 mg/mL, Sigma P5380), EDTA (5 mM, Sigma 03690), EGTA (5 mM, BioWorld, 40520008-1), DNase I (10 µg/mL, Roche 11 284 932 001), and Accutase (0.68X, Sigma, A6964) in a total volume 20 µL per reaction. For single organoid dissociation, Nunc MicroWell plates (Nunc, 438733) and siliconized p10 pipette tips (VWR, 53509-134) were used.

Split organoid experiment

Organoids for the split organoid experiment were cultured in ENR medium as previously described.[20] Single organoids, derived from Days 2 – 6 post crypt splitting were isolated from Matrigel as described above. Subsequently, single organoids were isolated by hand-picking into a 384 well plate containing single organoid dissociation mix. As before, single organoids were dissociated by combining triaturation using non-filter pipette tips every 5 minutes and incubation at 37°C for 15 minutes in 100 µL volume. Finally, the

dissociation mix was diluted with cell buffer. The single cell suspension of one organoid was split into two separate samples, and introduced subsequently on the system.

Material and reagent list for the split organoid experiment: Intestinal organoids were cultured in Matrigel (Corning, 356230) with organoid base medium supplemented with ENR (not containing CV). Organoid base medium was prepared using DMEM/F12 (Gibco, 11320033), Hepes (100 mM, Gibco, 15630056), penicillin-streptomycin (100 U/mL, Gibco, 15140122), B27 supplement (1 µM, Gibco, 17504-044), N2 supplement (1 µM, Gibco, 17502001), and N-Acetyl-L-cysteine (1 µM, Sigma, A9165). ENR medium was prepared using base medium (as above), EGF (E, 50 ng/mL, LifeTechnologies, PMG8043), mNoggin (N, 100 ng/mL, produced in-house), R-spondin (R, 1 µg/mL, produced in-house).

Single cell isolation from the small mouse intestine

Crypts were isolated from single small intestines of 7-week old male C57BL/6J mice following the protocol from Bas and Augenlicht *et al.*[17]. Briefly, the small intestine of a single mouse was isolated and then washed both on the inside and outside with ice cold PBS. The small intestine was cut open longitudinally and washed again with PBS. The intestine was then digested non-enzymatically for 3 minutes in PBS/EDTA/DTT. Next, the tissue was cut into small pieces and transferred into a 50 mL Falcon tube containing 20 mL of ice-cold PBS. The PBS solution containing the tissue pieces was gently triturated 10 times using a 10 mL pipette. After tissue fragments sedimented, the supernatant was removed, and the process was repeated three more times until the supernatant was clear. Next, the supernatant was removed, PBS/EDTA was added and the sample incubated for 30 minutes at 4°C on a rocking plate. After incubation, tissue fragments were left for sedimentation (up to 5 minutes), then the supernatant was removed. Subsequently, tissue fragments were triturated with ice cold PBS by pipetting up and down. After large tissue fragments sedimented (up to 5 minutes), the supernatant containing crypts was collected as Fraction 1 (F1). Fraction collection was repeated five times (F2-F5), followed by trituration with ice cold PBS, while each fraction was stored separately. Each fraction was inspected for cell debris and villus contamination.

For single cell bulk sample preparation, crypts from F2 or F3 were spun down at 600 x g for 10 minutes (brake 5). Following centrifugation, the supernatant was removed and cells were enzymatically

dissociated for 1 minute at 37°C. Cells were then washed two times in PBS/BSA and strained two times using a Flowmi 40 µm strainer to minimize the amount of multiplets. Cell suspensions were diluted in cell-loading buffer and loaded on the DisCo chip.

For single cell isolation from single crypts, crypts from F3 were transferred to FBS-coated 6 well plates. Subsequently, single crypts were isolated by hand-picking after which they were transferred to a Nunc microwell culture plate containing single crypt dissociation mix. Single crypts were dissociated by combining trituration using non-filter pipette tips every 5 minutes and incubation at 37°C for a total of 15 minutes. As also noted in the Results section, obtaining a true single cell suspension proved highly challenging, despite testing several dissociation buffer compositions (**Supplementary Table 3**), given that many cells were lost or were only partially recovered in multiplets / clumps. Following dissociation, cell suspensions were diluted in cell loading buffer and loaded on the DisCo chip.

Material and reagent list for single cell isolation from mouse intestinal crypts: For small intestine washing, PBS (Gibco, 14190-094) was used. For the non-enzymatic dissociation of small intestinal pieces, PBS (Gibco, 14190-094), EDTA (3 mM Sigma, 3690) and DTT (0.5 mM, Applichem, A2948,0005) was used. Intestinal pieces were incubated in PBS (Gibco, 14190-094), EDTA (2 mM, Sigma, 3690) followed by fraction collection in PBS (Gibco, 14190-094). Bulk single cell crypt preparations from the small intestine were prepared using PBS (Gibco, 14190-094), TrypLE select (1X A1217701, Gibco) and DNase I (10 mg/mL, Roche 11 284 932 001) in a total volume of 500 µL per reaction. Cells were then washed using PBS (Gibco, 14190-094), BSA (0.01%, Sigma, B8667) and strained using a Flowmi 40 µm (Sigma, BAH136800040-50EA) strainer into 500 µL final volume. Single cells were dissociated from single crypts using PBS (Gibco, 14190-094), *B. licheniformis* protease (20 mg/mL, Sigma P5380), EDTA (10 mM, Sigma 03690), EGTA (10 mM, BioWorld, 40520008-1), DNase I (20 µg/mL, Roche 11 284 932 001), and Accutase (0.6X, Thermofischer, A1110501) in a total volume of 20 µL per reaction. For single crypt dissociation, Nunc MicroWell plates (Nunc, 438733) and non-filtered 10 µL pipette tips (VWR, 53509-134) were used.

RNA Fluorescence in situ hybridization (RNAscope) on intestinal organoids

For the RNAscope assay, organoids in matrigel were fixed in PFA at 4°C overnight. The next day, organoids were washed with PBS and embedded in histogel. Histogel blocks were subsequently infiltrated with

paraffin using a standard histological procedure (VIP6, Sakura). RNAscope Multiplex Fluorescent V2 assay was performed according to the manufacturer's protocol on 4 µm paraffin sections, hybridized with the probes Mm-Ly6a-C2, Mm-Fabp1-C1, Mm-Muc2-C2, Mm-PpiB-C2 positive control, and Duplex negative control at 40°C for 2 hours and revealed with TSA Opal650 for C1 channel and TSA Opal570 for C2 channel. Tissues were counterstained with DAPI and mounted with Prolong Diamond Antifade Mountant. Slides were imaged on an Olympus VS120 whole slide scanner (Olympus). The resulting images were converted to the TIFF file format using the Fiji (version 1.52p) plugin BIOP VSI Reader (version 7). ROIs were extracted using a custom Python (version 2.7.15) script and the PIL library (version 6.2.2). Brightness of the extracted ROIs was adjusted in Fiji: Images of one target were loaded, stacked, brightness adjusted for the whole stack using the *setMinAndMax()* function. Finally, images were unstacked, merged with other channels, and exported as PNG files. Materials and reagents are listed in the Material and reagent list, points 17 - 18.

Sequencing, analysis, barcode correction

The data analysis was performed using the Drop-seq tools package (version 2.3.0, https://github.com/broadinstitute/Drop-seq/releases/tag/v2.3.0)[3,15] on the EPFL SCITAS HPC platform. All data pre-processing steps were done according to the Drop-seq tools manual, except for the *DetectBeadSubstitutionErrors* function, which was not utilized and replaced by the barcode merging strategy described below. After trimming and sequence tagging, reads were aligned to the human (hg38), mouse (GRCm38), or mixed reference genomes[3] (GSE63269), depending on the origin of the cellular input material, using STAR (version 2.7.0.e)[18]. Following alignment, BAM files were processed to obtain initial read-count matrices (RCM) per sample (Note: DGE summary files were used for experiments displayed in **Figure 1H** and **Figure 1I**). Cell barcodes were prefiltered at > 35 UMIs (for the species mixing experiment, the sum of 35 UMIs for both species was used as a prefiltering criterion). Graphs were built by identifying barcodes connected by Levenshtein distance 1. For each graph, the barcode containing the highest number of UMIs was identified as the central barcode. The graphs were pruned (barcodes removed) at a Levenshtein distance > 2 to the central barcode, the remaining barcodes in the graph were merged.

For cell recovery efficiency experiments using the DISPENCELL platform (**Figure 1I**) and for Drop-seq comparison experiments (**Figure 1J**) barcodes encompassing at least 500 UMIs were compiled into the RCMs. Additionally, prior to Drop-seq comparison experiments, processed BAM files were down sampled to the same read depth using samtools (http://www.htslib.org/doc/samtools.html). Box plot elements depicting UMI counts per cell (**Figure 1I**) represent the following values: centerline, median; box limits, upper/lower quartiles; whiskers, 1.5x interquartile range; points, UMIs per cell.

Time course organoid kinetic analysis

RCMs were further processed via R (version 3.6.2) using *Seurat* (version 3.1.1) and *uwot* (version 0.1.3)[19]. Per individual organoid-RCM cells with > 800 features, < 7.5% mitochondrial reads were retained in the analysis. The time course kinetics of organoids were processed in three independent experiments, which were considered as three individual batches. The three independent experiments were merged using *FindIntegrationAnchors(list(experimental_batches), anchor.features = 80, dims = 1:12, k.filter = 200, k.anchor = 8)* and *IntegrateData().* Data was scaled and PCAs computed using default settings. Uniform Manifold Approximation and Projection (UMAP) dimensional reduction via *RunUMAP()* and *FindNeighbors()* were performed using the first 12 PCA dimensions as input features. *FindClusters()* was computed at resolution 0.75.

The intestinal organoids for the split organoid experiment were processed in four independent experiments, which were considered as four individual batches, each encompassing at least two independent single intestinal organoids. The four independent experiments were merged using *FindIntegrationAnchors(list(experimental_batches), anchor.features = 120, dims = 1:10, k.filter = 100, k.anchor = 12)* and *IntegrateData().* Data was scaled and PCAs computed using default settings. Uniform Manifold Approximation and Projection (UMAP) dimensional reduction via *RunUMAP()* and *FindNeighbors()* was performed using the first 14 PCA dimensions as input features. *FindClusters()* was computed at resolution 0.9.

The intestinal crypts were processed in five independent experiments, which were considered as five individual batches each encompassing single intestinal crypts and pooled (bulk) samples. The five independent experiments were merged using *FindIntegrationAnchors(list(experimental_batches),*

anchor.features = 150, dims = 1:10, k.filter = 150, k.anchor = 10) and *IntegrateData().* Data was scaled and PCAs computed using default settings. Uniform Manifold Approximation and Projection (UMAP) dimensional reduction via *RunUMAP()* and *FindNeighbors()* was performed using the first 15 PCA dimensions as input features. *FindClusters()* was computed at resolution 0.8.

Combined intestinal crypts and organoids were processed as eight independent batches. These eight batches were merged using *FindIntegrationAnchors(list(experimental_batches), anchor.features = 150, dims = 1:15, k.filter = 150, k.anchor = 10)* and *IntegrateData().* Data was scaled and PCAs computed using default settings. Uniform Manifold Approximation and Projection (UMAP) dimensional reduction via *RunUMAP()* and *FindNeighbors()* was performed using the first 15 PCA dimensions as input features. *FindClusters()* was computed at resolution 0.9. Merging retained the global grouping of the data but introduced minor annotation discrepancies in similar clusters between the individual and merged datasets. For example cells that were annotated TA-G1 in the crypt data (Supplementary Figure 4D) were annotated as stem cells in the merged data (Supplementary Figure 5D).

Merged data was visualized using the Seurat intrinsic functions *VlnPlot()*, *FeaturePlot()*, *DotPlot(), DimPlot().* Differentially expressed genes per cluster were identified using *FindAllMarkers()* using default parameters. The Seurat-Object is accessible via GSE148093. Cumulative Z-scores were calculated based on the scaled expression per cell across the defined gene signatures[20,21]. Pie-chart, bubble-plot and bargraph visualizations were carried out with *ggplot2*.

C1 HEK library processing and analysis

Libraries were sequenced on a NextSeq500 sequencer (Illumina) in paired-end run format (read 1 - 16 bp, read 2 - 59 bp) with an average of 3 x 10^6 reads per library. The read quality of sequenced libraries was evaluated with *FastQC*. Sequencing reads were aligned to the reference human genome assembly GRCh38.90 using STAR[18]. Reads aligned to annotated genes were quantified with *htseq-count* [22].

Slingshot analysis

The trajectories were constructed using the Slingshot wrapper implemented in the dyno package (https://github.com/dynverse/dyno)[23]. The method was provided with the first 5 dimensions of a multi-

dimensional scaling as dimensionality reduction, the clustering as described earlier, and the stem cell cluster as starting cell population. All other parameters were left at default settings. Genes that change along the trajectory were ranked using the *calculate_overall_feature_importance* function from the dynfeature package (version 1.0, https://github.com/dynverse/dynfeature), and the top 50 differentially expressed genes were selected. The dynplot package (version 1.1, https://github.com/dynverse/dynplot) was used to plot the trajectory within a scatterplot and heatmap.

Psupertime analysis

Cell labels and sample-day labels were extracted from the merged and batch-corrected meta-data of the Seurat object to run *psupertime,* a method of identifying genes relevant to biological processes using cell-level temporal labels to build a l1 regularised ordinal logistic regression model (Macnair & Claassen, biorxiv 2019)[24]. Sample-day labels indicating the experimental temporal order were used to conduct a *psupertime* analysis on batch-corrected and normalized gene expression data of cells, with selected cell type labels. The analysis was performed including all genes and encompassing a 10-fold cross-validation using default settings. Genes with coefficients (beta-values) greater than zero were considered relevant for the temporal expression dynamics. Expression of relevant genes was plotted per organoid per cell.

Material and reagent list for all experiments

Material information is listed in the following format: Material name (vendor, ordering number). Reagent information is listed in the following format: Reagent name (final concentration in the solution, vendor, order number).

1. For microfluidic device fabrication SU8 3010 (Microchem) negative photoresist, AZ40XT (Microchem) positive photoresist, HFE-7500 (3M, Novec 297730-93-9), Trichloro(1H, 1H, 2H, 2H - perfluorooctyl) silane (1%, Aldrich, 448931), and biopsy punchers (Darwin microfluidics, KPUNCH05) were used.
2. For microfluidic device handling Prot/Elec 200 µL gel loading tips (Biorad, #223-9915), dH_2O (Invitrogen, 10977035), tygon tubing (Cole-Parmer, GZ-06420-02), beads (Chemgenes, lot

051917, Macosko-2011-10), droplet generation oil (Biorad, 186-4006), murine RNase inhibitor (100 U, NEB, M0314L) were used. Cell wash buffer was prepared using PBS (1X, Gibco, 14190-094) and BSA (0.01%, Sigma, B8667). Cell loading buffer was prepared using PBS (1X, Gibco, 14190-094), Optiprep (6%, Sigma, D1556), and BSA (0.01%, Sigma, B8667). Lysis buffer was prepared from Optiprep (28%, Sigma, D1556), Sarkosyl (2.2%, Sigma, L7414), EDTA (20 mM, Sigma, 3690), Tris (100 mM, Sigma, T2944), DTT (50 mM, Applichem, A2948,0005).

3. For sample washing prior to reverse transcription, SSC (6X, Sigma, S6639) and dH_2O (Invitrogen 10977-035) were used.
4. For reverse transcription (RT) reaction dH_2O (Invitrogen, 10977-035), Ficoll PM-400 (4%, Sigma, F5415), dNTPs (1mM, Thermo, R0193), murine RNase inhibitor (100U, NEB, M0314L), Maxima H- reverse transcriptase (500 U, Thermo Scientific, EP0753), Template Switching Oligo (AAGCAGTGGTATCAACGCAGAGTGAATrGrGrG, 2.5 µM, IDT) were used in a total volume 50 µL per reaction.
5. For exonuclease I reaction exonuclease I (100 U, NEB, M0293L) and exonuclease buffer were used in a total volume 50 µL per reaction.
6. For cDNA amplification Kapa HiFi Hot start ready mix 2X (Roche, KK2602), dH_2O (Invitrogen, 10977035), and SMART PCR primer (AAGCAGTGGTATCAACGCAGAGT, 0.8 µM, IDT) used in a total volume 50 µL per reaction. CleanPCR magnetic beads (0.6X ratio, GC biotech, CPCR-0050), Fragment Analyzer (Agilent, DNF-474-0500 kit), and Qubit HS sensitivity kit (Invitrogen, Q33231) were used for cDNA purification and quantification.
7. For library preparation in-house produced Tn5 was used. To stop tagmentation, SDS was used (0.2%, Sigma, 71736). For library amplification Kapa HiFi kit with dNTPs (Roche, KK2102), P5 SMART PCR (AATGATACGGCGACCACCGAGATCTACACGCCTGTCCGCGGAAGCAGTGGTATCAA CGCAGAGT*A*C, 0.3 µM, IDT), custom Nextera oligos[25] (0.3 µM, IDT) and dH_2O (Invitrogen, 10977035) were used. Libraries were purified and quantified using CleanPCR magnetic beads (0.6X ratio, GC biotech, CPCR-0050), Fragment Analyzer (Agilent, DNF-474-0500 kit), and Qubit HS sensitivity kit (Invitrogen, Q33231).

8. TE-TW wash buffer was prepared in dH_2O (Invitrogen, 10977035) using Tris (10 mM, Sigma T2944), EDTA (1mM, Sigma, 3690), and Tween 20 (0.01%, Sigma, P9416).
9. TE-SDS wash buffer was prepared in dH_2O (Invitrogen, 10977035) using Tris (10 mM, Sigma, T2944), EDTA (1 mM, Sigma, 03690), and SDS (0.5%, Sigma, 71736).
10. Tris wash buffer was prepared in dH_2O (Invitrogen, 10977035) using Tris (10 mM, Sigma, T2944).
11. For mammalian cell culture dissociation and counting Trypsin-EDTA (Gibco, 25200056) and trypan blue were used (0.4%, Thermo Fisher Scientific, T10282). Cell culture medium was prepared using DMEM Glutamax (Gibco, 10565018), FBS (10%, Gibco, 10270106) and penicillin-streptomycin (100 U/mL, Gibco, 15140122). Cell wash and cell loading buffers were prepared as described above.
12. Intestinal organoids were cultured in Matrigel (Corning, 356230) with organoid base medium (described in point 13) supplemented with ENR (+ CV where indicated) and rock inhibitor (where indicated, Sigma, Y0503).
13. Organoid base medium was prepared using DMEM/F12 (Gibco, 11320033), Hepes (100 mM, Gibco, 15630056), penicillin-streptomycin (100 U/mL, Gibco, 15140122), B27 supplement (1 µM, Gibco, 17504-044), N2 supplement (1 µM, Gibco, 17502001), and N-Acetyl-L-cysteine (1 µM, Sigma, A9165).
14. ENR medium was prepared using base medium (as above), EGF (E, 50 ng/mL, LifeTechnologies, PMG8043), mNoggin (N, 100 ng/mL, produced in-house), R-spondin (R, 1 µg/mL, produced in-house).
15. ENR CV medium was prepared with addition of CHIR (C, 3 µM, CalBiochem, CHIR99021), and Valproic acid (V, 3 mM, Sigma P4543) to ENR medium.
16. Single-organoid single-cell dissociation mix was prepared using PBS (Gibco, 14190-094), *B. licheniformis* protease (10 mg/mL, Sigma P5380), EDTA (5 mM, Sigma 03690), EGTA (5 mM, BioWorld, 40520008-1), DNase I (10 µg/mL, Roche 11 284 932 001), and Accutase (0.68X, Sigma, A6964) in a total volume 20 µL per reaction. For single organoid dissociation Nunc MicroWell plates (Nunc, 438733) and siliconized p10 pipette tips (VWR, 53509-134) were used.

17. For intestinal organoid preparation for RNAscope, cold Cell Recovery Solution (Corning, 354253), Histogel (Thermo Scientific, HG-4000-012), Paraformaldehyde (4%, PFA, Electron Microscopy Sciences, 15714) were used.
18. For the RNAscope assay, organoids were stained using RNAscope Multiplex Fluorescent V2 assay (ACD Bio-Techne, 323110), Ly6a probe (ACD Bio-Techne, 427571-C2), Fabp1 probe (ACD Bio-Techne, 562831), Muc2 probe (ACD Bio-Techne, 315451-C2), PpiB probe (ACD Bio-Techne, 313911-C2), Duplex negative control (ACD Bio-Techne, 320751), TSA Opal650 (Perkin Elmer, FP1496001KT), TSA Opal570 (Perkin Elmer, FP1488001KT), and Prolong Diamond Antifade Mountant (Thermo Fisher, P36965).

Data availability

The GEO accession number for scRNA-seq data reported in this paper is GSE148093. For reviewing purposes, the temporary access token is: ejwxsgekplwdzwv.

The raw data and count matrices for **Figure 1H** and **Supplementary Figure 1H** are stored under the access code GSM4454017. The raw data and count matrices for **Figure 1I** and **Supplementary Figure 1F** are available under the access code GSM4454017. The raw data and count matrices for **Figure 1J** are stored under the access codes GSM4454012 - GSM4454016. The raw data and count matrices for **Supplementary Figure 1J&K** are stored under the access code GSM5567775 - GSM5567779. The raw data and count matrices for **Supplementary Figure 1L** are stored under the access codes GSM5567571 - GSM5567730. The raw data and count matrices for **Supplementary Figure 2F&G** are stored under the access codes GSM5567845 - GSM5567854. The raw data for intestinal organoids embedded in **Figure 2**, **Supplementary Figure 2**, **Figure 3A&E&J&K**, **Supplementary Figure 3C&D** and **Supplementary Figure 5A-D** are stored under access codes GSM4453981- GSM4454011. The raw data and count matrices for intestinal crypts embedded in **Figure 3G-K**, **Supplementary Figure 4A-D** and **Supplementary Figure 5A-D** are stored under the access codes GSM5567818 - GSM5567844.

Code availability

This technology has been developed as an open source platform, so all required know-how for its implementation, e.g. the custom machine-vision code and barcode merging script, will therefore be rendered publicly available upon publication. The source code for the machine-vision code is already available on github (https://github.com/DeplanckeLab/DisCo_source).

Supplementary references

1. Zheng, G. X. Y. *et al.* Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8**, 1–12 (2016).

2. Klein, A. M. *et al.* Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* **161**, 1187–1201 (2015).

3. Macosko, E. Z. *et al.* Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161**, 1202–1214 (2015).

4. Zhang, X. *et al.* Comparative Analysis of Droplet-Based Ultra-High-Throughput Single-Cell RNA-Seq Systems. *Mol. Cell* 1–13 (2018). doi:10.1016/j.molcel.2018.10.020

5. Picelli, S. *et al.* Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protoc.* **9**, 171–81 (2014).

6. Hashimshony, T. *et al.* CEL-Seq2: sensitive highly-multiplexed single-cell RNA-Seq. *Genome Biol.* **17**, 77 (2016).

7. Hwang, B., Lee, J. H. & Bang, D. Single-cell RNA sequencing technologies and bioinformatics pipelines. *Exp. Mol. Med.* **50**, 1–14 (2018).

8. Pollen, A. A. *et al.* Low-coverage single-cell mRNA sequencing reveals cellular heterogeneity and activated signaling pathways in developing cerebral cortex. *Nat. Biotechnol.* **32**, 1053–1058 (2014).

9. Goldstein, L. D. *et al.* Massively parallel nanowell-based single-cell gene expression profiling. *BMC Genomics* **18**, 1–10 (2017).

10. Gierahn, T. M. *et al.* RNA sequencing of single cells at high throughput. *Nat. Methods* **14**, 395–398 (2017).

11. Wang, Y. *et al.* Comparative analysis of commercially available single-cell RNA sequencing platforms for their performance in complex human tissues. *bioRxiv* 541433 (2019). doi:10.1101/541433

12. Ding, J. *et al.* Systematic comparison of single-cell and single-nucleus RNA-sequencing methods. *Nat. Biotechnol.* **38**, 737–746 (2020).

13. Biocanin, M., Bues, J., Dainese, R., Amstad, E. & Deplancke, B. Simplified Drop-seq workflow with minimized bead loss using a bead capture and processing microfluidic chip. *LoC* **19**, 1610–1620 (2019).

14. Picelli, S. *et al.* Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. *Genome Res.* **24**, 2033–2040 (2014).

15. Macosko, E., Goldman, M. & McCarroll, S. Drop-Seq Laboratory Protocol version 3.1. 1–20 (2015). Available at: http://mccarrolllab.org/download/905/.

16. Gjorevski, N. *et al.* Designer matrices for intestinal stem cell and organoid culture. *Nature* **539**, 560–564 (2016).

17. Bas, T. & Augenlicht, L. H. Real time analysis of metabolic profile in Ex vivo mouse intestinal crypt organoid cultures. *J. Vis. Exp.* **e52026**, 1–5 (2014).

18. Dobin, A. *et al.* STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

19. Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888–1902 (2019).

20. Serra, D. *et al.* Self-organization and symmetry breaking in intestinal organoid development. *Nature* **562**, 66–72 (2019).

21. Gregorieff, A., Liu, Y., Inanlou, M. R., Khomchuk, Y. & Wrana, J. L. Yap-dependent reprogramming of Lgr5+ stem cells drives intestinal regeneration and cancer. *Nature* **526**, 715–718 (2015).

22. Anders, S., Pyl, P. T. & Huber, W. HTSeq-A Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).

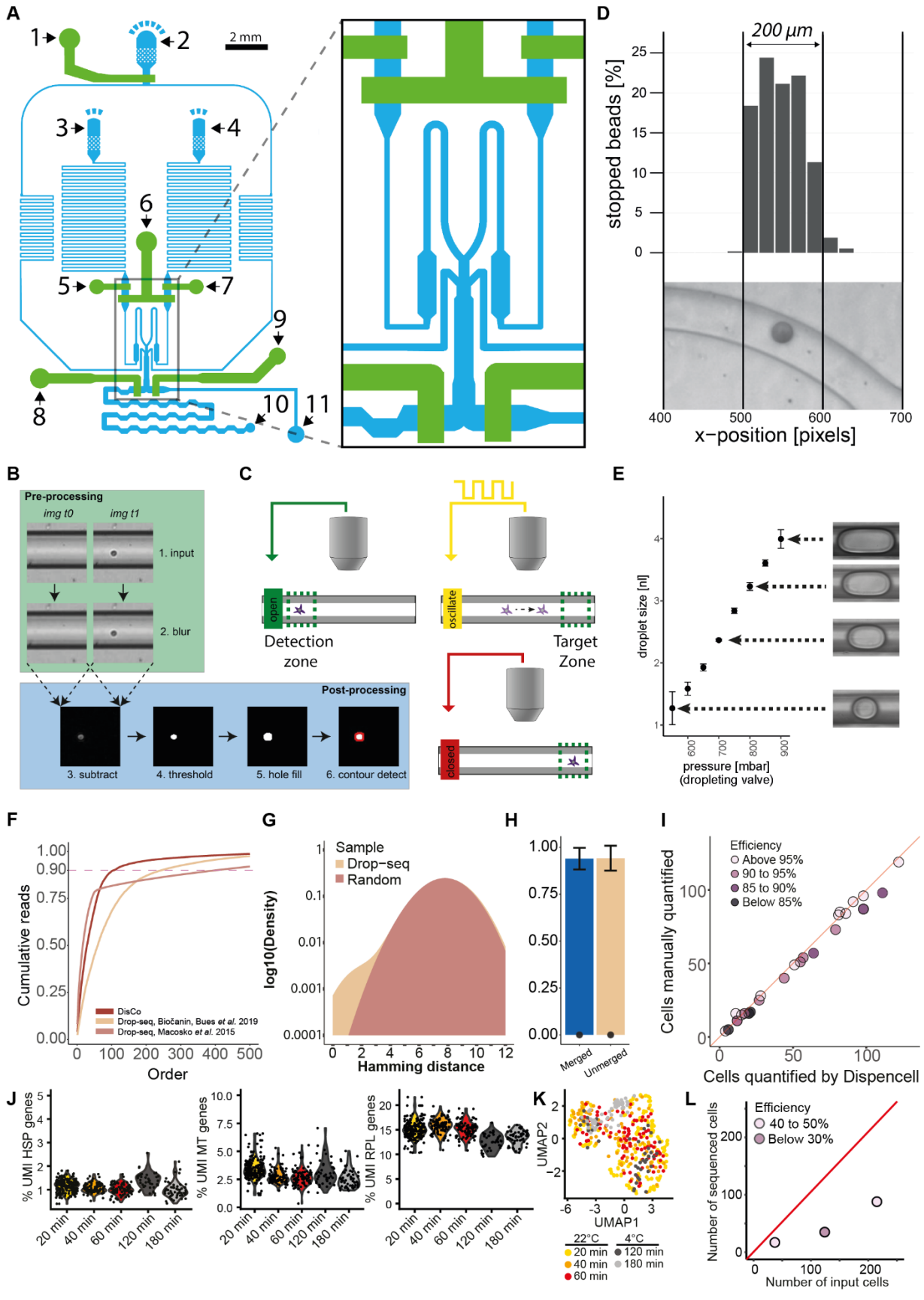
23. Saelens, W., Cannoodt, R., Todorov, H. & Saeys, Y. A comparison of single-cell trajectory inference methods. *Nat. Biotechnol.* **37**, 547–554 (2019).

24. Macnair, W. & Claassen, M. psupertime: supervised pseudotime inference for single cell RNA-seq data with sequential labels. *bioRxiv* 622001 (2019). doi:10.1101/622001

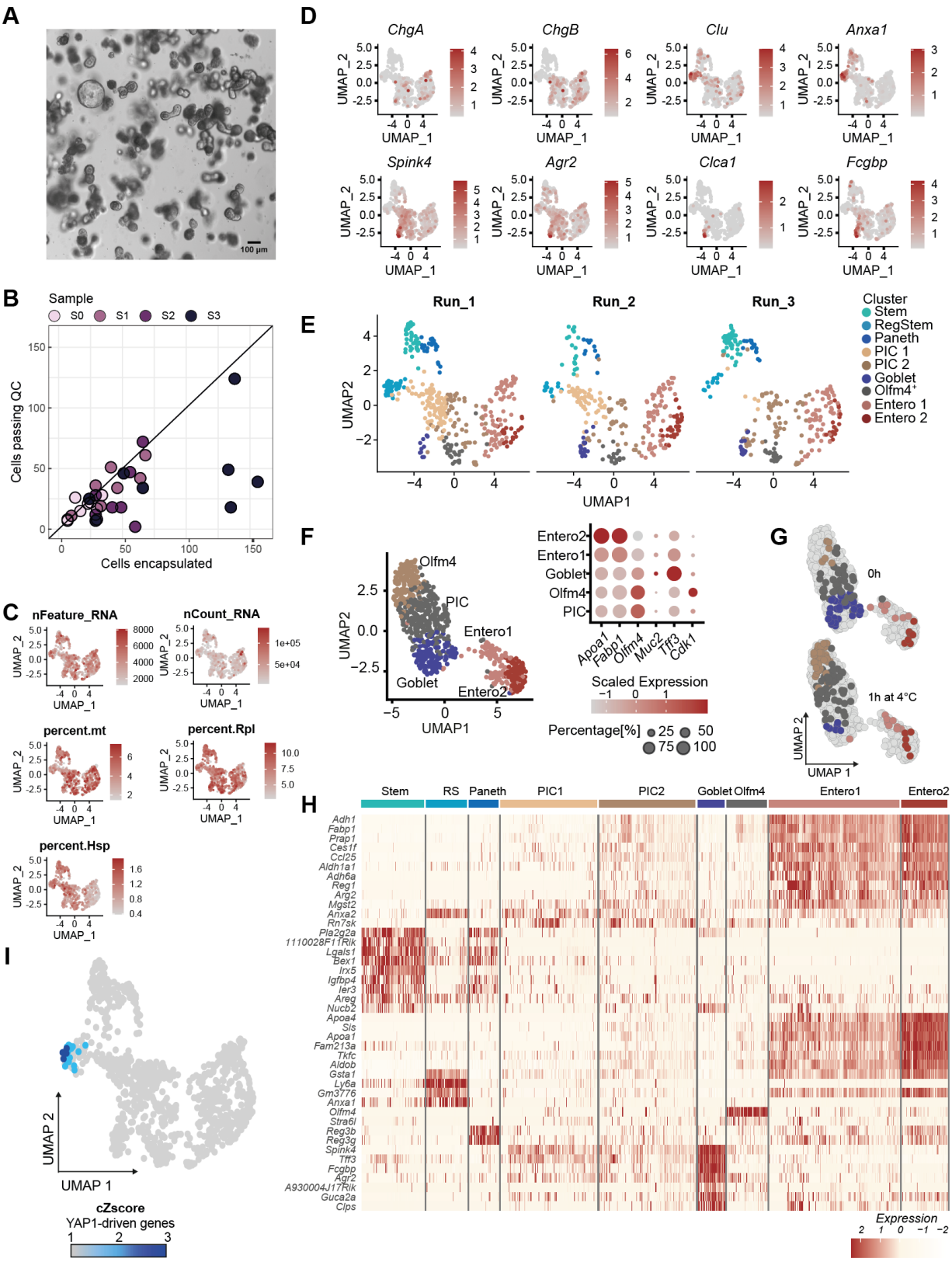
25. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013).

26. Barker, N., Van Oudenaarden, A. & Clevers, H. Identifying the stem cell of the intestinal crypt: Strategies and pitfalls. *Cell Stem Cell* **11**, 452–460 (2012).

2. Supplementary Figures

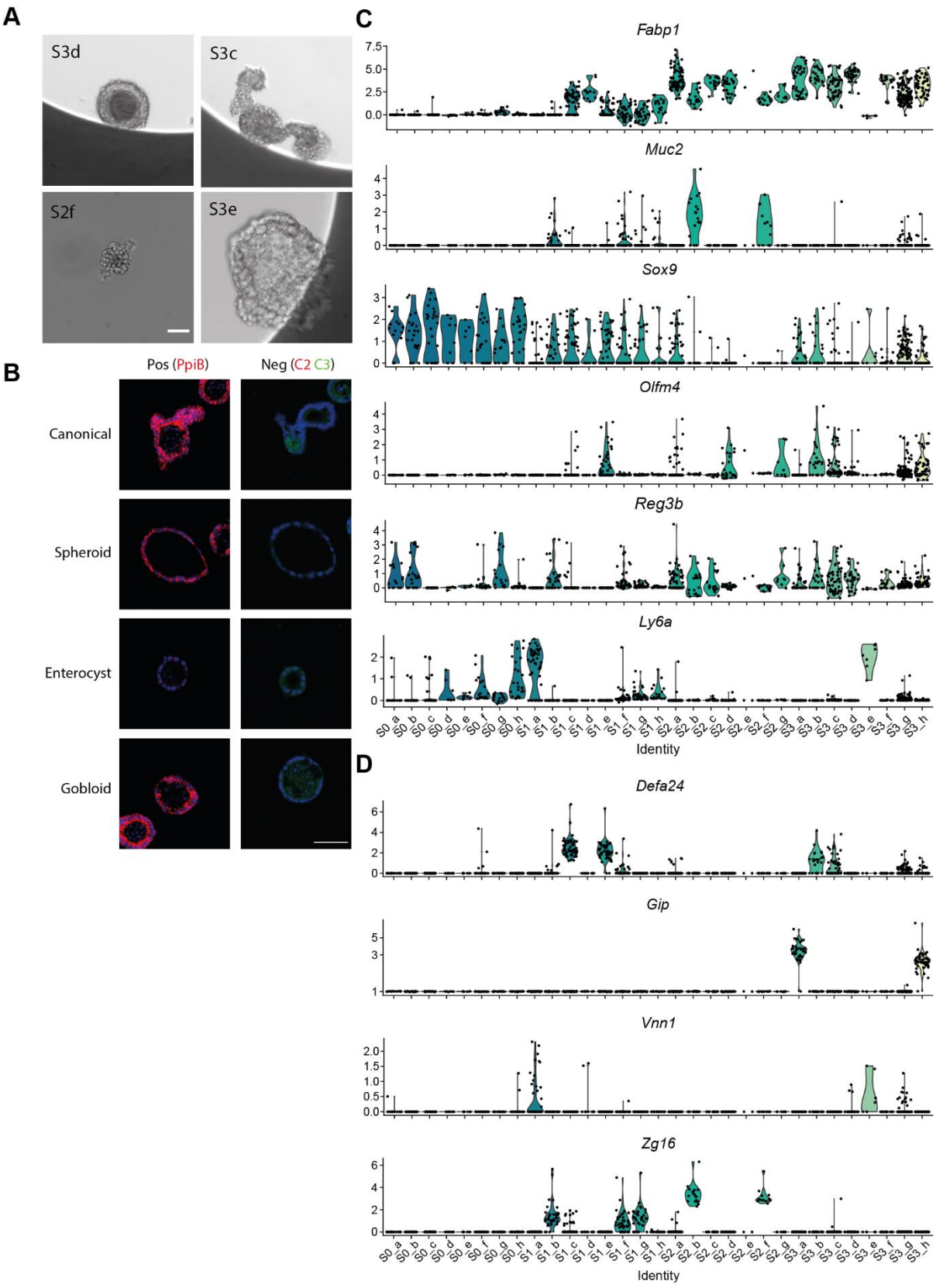

A
1
2
2 mm
3
4
5
6
7
8
9
10
11
D
200 µm
stopped beads [%]
x−position [pixels]
B
Pre-processing
img t0
img t1
1. input
2. blur
Post-processing
3. subtract
4. threshold
5. hole fill
6. contour detect
C
open
oscillate
closed
Detection zone
Target Zone
E
droplet size [nl]
pressure [mbar] (dropleting valve)
F
Cumulative reads
Order
DisCo
Drop-seq, Bičanin, Bues et al. 2019
Drop-seq, Macosko et al. 2015
G
Sample
Drop−seq
Random
log10(Density)
Hamming distance
H
Merged
Unmerged
I
Efficiency
Above 95%
90 to 95%
85 to 90%
Below 85%
Cells manually quantified
Cells quantified by Dispencell
J
% UMI HSP genes
% UMI MT genes
% UMI RPL genes
20 min
40 min
60 min
120 min
180 min
K
UMAP1
UMAP2
22°C
4°C
20 min
40 min
60 min
120 min
180 min
L
Efficiency
40 to 50%
Below 30%
Number of sequenced cells
Number of input cells

Supplementary Figure 1: (**A**) Schematic of the DisCo device design (blue: flow layer, green: control layer). 1. oil valve, 2. oil inlet, 3. cell inlet, 4. bead inlet, 5. cell valve, 6. dropleting valve, 7. bead valve, 8. sample valve, 9. waste valve, 10. sample outlet, 11. waste outlet. (**B**) Real-time image processing for particle detection. Two consecutive images are de-speckled by Gaussian blurring, and subtracted. The resulting image is thresholded and holes are filled by dilatation. Finally, contours are detected and classified by size and circularity thresholding. (**C**) Particle positioning by valve oscillation. Approaching particles are detected in the detection zone. Once a particle is detected, the channel valve is oscillated to induce discrete movements of particles. Oscillation is terminated once correct placement of a particle is achieved. (**D**) Stopping accuracy in a defined window. Beads (n = 744) were positioned using valve oscillation, their position was manually determined within the stopping area. Scale was approximated from channel width. (**E**) Volume-defined droplet on-demand generation by valve pressurization. Droplets (n = 68, ~8 per condition) were produced by pressurizing the dropleting valve at different pressures. Size was determined by imaging the dropleting process. Volumes were calculated from the imaging data based on droplet length and channel geometry. Thus, they should be considered an approximation. Error bars represent the standard deviation. The channel width of displayed images is 250 µm. (**F**) Cumulative reads per barcode (n = 500) for DisCo and two Drop-seq experiments[3,13]. (**G**) Hamming distances between all 12 nt barcodes of a Drop-seq experiment, and generated 12 nt random barcode sequences representing the probability density for each set of barcodes. (**H**) Species purity (bars) and doublet ratio (dots) for unmerged and merged barcodes. Error bars represent the standard deviation. (**I**) Correlation of the number of manually counted cells by fluorescence microscopy and the number of cells quantified by the DISPENCELL platform. (**J-K**) HEK 293T cells were processed with DisCo at 22ºC after 20, 40 or 60 min or stored on ice for 120 or 180 min and subsequently processed. (**J**) Violin plots showing the percentage of UMIs per cell of heat-shock-protein (HSP), mitochondrial protein-coding (MT), or ribosomal protein-coding (RPL) genes. (**K**) UMAP embedding of all profiled HEK 293T cells from the five sampling time points, color-coded by sampling time. (**L**) A quantified number of HEK 293T cells was processed with the Fluidigm C1 system. Processing efficiency was calculated as the percentage of cells retrieved from the sequencing data respective to the quantified number of input cells. The red line represents 100% efficiency, and samples were colored according to the recovery efficiency after sequencing.

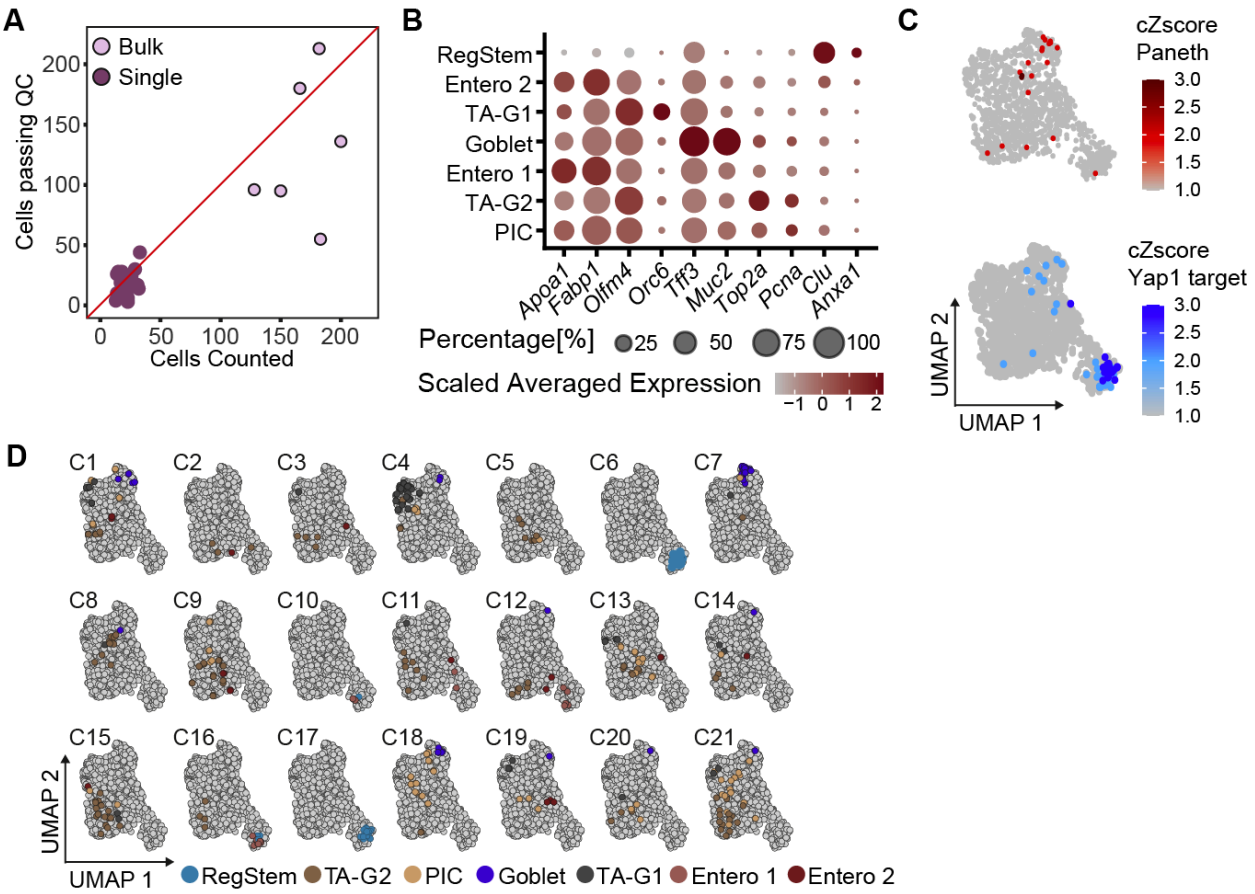


Supplementary Figure 2: (A) Representative brightfield image of a differentiated organoid culture from single LGR5⁺ cells. (**B)** Correlation of encapsulated cells on-chip with the number of cells detected after sequencing (cells passing QC, filtered above 800 genes/cell). (**C**) UMAP embedding colored by the number

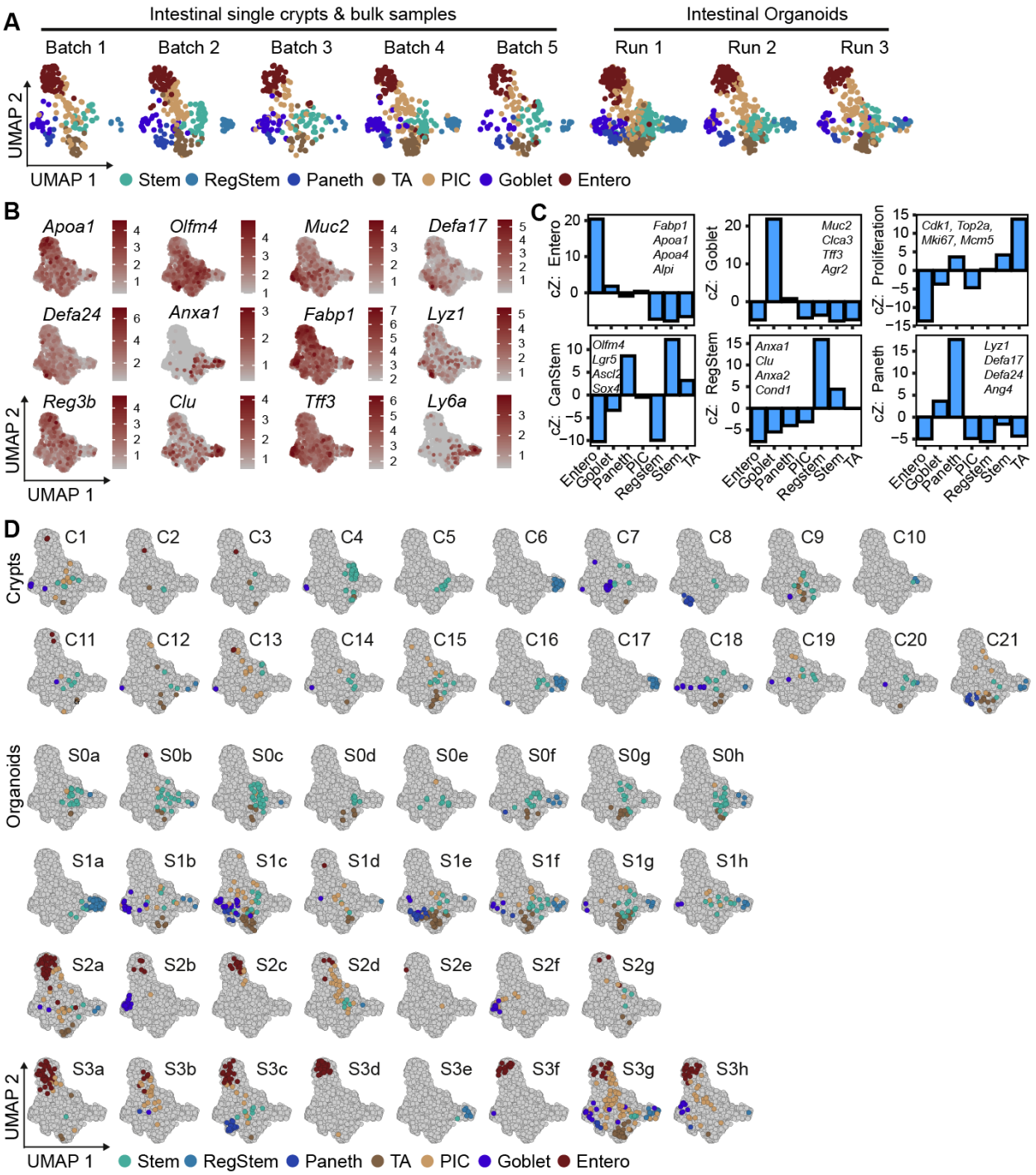
of detected UMIs per cell, the number of detected genes per cell, the percentage of mitochondrial reads, and the percentage of reads mapping to genes coding for respectively ribosomal proteins (Rpl), and heat-shock proteins (Hsp). (**D**) UMAP embedding colored by expression of selected marker genes (*Clu*, *Anxa1*, *Spink4*, *ChgB*, *ChgA*, *Agr2*, *Clca1*, and *Fcgbp*). (**E**) UMAP embedding for each of the three independent experimental batches colored by cluster annotation. (**F**) UMAP embedding of cells collected from nine additional individual organoids (under maintenance conditions) for the purpose of evaluating batch effects. *Left:* All 748 processed cells clustered with k-means clustering, after which clusters were annotated according to marker gene expression. *Right:* Expression dot plot of selected marker genes. (**G**) Projection of cells (colored by cell type) derived from one organoid that was split into two independent samples (split organoid) on the reference UMAP shown in F). Organoid "S2_2" was split into two batches, which were processed subsequently, with a one-hour delay, during which the second batch was stored at 4ºC. (**H**) Heatmap of top DE genes per annotated cluster. (**G**) YAP1 target gene activity on a UMAP embedding. The expression of genes that are positively regulated by YAP1[21] was calculated as the cumulative Z-score and projected on the UMAP embedding of all sequenced cells.


A
S3d
S3c
S2f
S3e
B
Pos (PpiB)
Neg (C2 C3)
Canonical
Spheroid
Enterocyst
Gobloid
C
Fabp1
Muc2
Sox9
Olfm4
Reg3b
Ly6a
Identity
D
Defa24
Gip
Vnn1
Zg16
Identity

Supplementary Figure 3: (**A**) Selected organoids imaged in microwell plates before dissociation to single cells. Scale bar: 50 µm. (**B**) RNAscope controls for organoids shown in **Figure 3C**. Positive control (*PpiB*), and negative control (Duplex negative). Scale bar: 50 µm. (**C**) Violin plots showing marker gene expression (*Fabp1, Muc2, Olfm4, Sox9, Reg3b, Ly6a*) per organoid. (**D**) Violin plots showing the expression of selected genes (*Defa24, Gip, Vnn1, Zg16*) identified via psupertime analysis per individual organoid.



Supplementary Figure 4: (**A**) Processing efficiency of DisCo for individual and bulk intestinal crypts. All cells processed with DisCo were manually counted during the experiment, and compared to cell numbers after quality filtering (> 500 UMIs). The red line represents 100% efficiency, and samples are colored according to sample type. (**B**) Expression dot plot of marker genes for clusters shown in **Figure 3G**. (**C**) Gene activity represented as the cumulative Z-score and projected on the UMAP embedding of all sequenced cells using the expression of *Top*: Paneth cell-associated genes encompassing *Lyz1*, *Defa17*, *Defa24* and *Ang4* and *Bottom*: genes that are positively regulated by YAP1.[21] (**D**) Projection of cell types onto the reference UMAP of cells derived from the 21 individual crypts. Cells per single crypt were colored according to their global clustering and highlighted on the UMAP embedding of all sequenced cells. Enterocytes (Entero), PIC (Potential intermediary cells), RegStem, (Regenerative Stem), TA (Transit amplifying cells; G1: G1/S and G2: G2/M cell cycle phase).



Supplementary Figure 5: (**A**) Combined UMAP embedding (as shown in **Figure 3J**) stratified by the five individual batches of intestinal crypt samples and the three independent experimental batches of intestinal organoid differentiation samples, collectively embedded and colored by cluster annotation. (**B**) UMAP-based visualization of the expression of specific markers that were used for cluster annotation. (**C**) Bar graph depicting the cumulative Z-score of the expression of genes that are indicated within the respective bar graph. *CanStem*: canonical stem cell, *RegStem*: regenerative stem cell. (**D**) Projection of cell types onto the reference UMAP of the *ex vivo* cell preparation for the 21 individual intestinal crypts and bulk samples

embedded together with the 31 individual intestinal organoids. Cells per single crypt or organoid are colored according to their global clustering and highlighted on the UMAP embedding of all sequenced cells.

3. Supplementary Tables 1-3

Supplementary Table 1: Cellular yield per intestinal organoid and intestinal crypt. *Top*: Summary of the number of cells per organoid that was obtained after sequencing and quality control. *Bottom*: Summary of the number of cells that was obtained per crypt (C) after sequencing and quality control (sequenced), 21 crypts were processed, split over five distinct experiments (batches). n.a. = not applicable.

Intestinal single organoids (sequenced)								
	a	**b**	**c**	**d**	**e**	**f**	**g**	**h**
S0	30	22	37	18	14	22	20	28
S1	61	36	72	14	58	44	39	22
S2	83	25	19	33	2	16	11	n.a.
S3	46	28	53	34	8	19	135	48

Intestinal single crypts (sequenced)					
Batch1	C1: 30	C2: 4	C3: 6		
Batch2	C4: 44	C5: 8	C6: 19	C7: 24	C8: 13
Batch3	C9: 26	C10: 3	C11: 11	C12: 14	
Batch4	C13: 28	C14: 8	C15: 22	C16: 17	C17: 17
Batch5	C18: 28	C19: 10	C20: 8	C21: 32	

Supplementary Table 2: DE genes for cell clusters (as shown in **Figure 2B**)

gene	p_val	avg_logFC	pct.1	pct.2	p_val_adj	cluster
Adh1	2.27631673711701e-67	1.17602093757581	0.991	0.639	1.82105338969361e-65	Entero1
Fabp1	8.97391864188553e-65	1.03782365836832	1	0.661	7.17913491350843e-63	Entero1
Apoa1	1.77992432836514e-63	0.805522082558538	0.981	0.557	1.42393946269211e-61	Entero1
Aldob	6.97360352458542e-60	0.694678513467298	1	0.782	5.57888281966834e-58	Entero1
Prap1	2.89992546219975e-59	0.819884545216953	0.995	0.752	2.3199403697598e-57	Entero1
Ces1f	3.67319503250653e-58	0.963767358595559	0.935	0.527	2.93855602600522e-56	Entero1
Ccl25	1.25004133755717e-55	0.901771305101302	0.986	0.693	1.00003307004574e-53	Entero1
Sis	1.20301185332912e-52	0.677799265079825	0.949	0.521	9.62409482663297e-51	Entero1
Aldh1a1	2.70977257321893e-50	0.885229974327549	0.94	0.656	2.16781805857514e-48	Entero1
Adh6a	1.98840147831558e-49	0.861278352992536	0.875	0.412	1.59072118265247e-47	Entero1
Reg1	9.99062930782402e-45	1.56303848520865	0.847	0.494	7.99250344625922e-43	Entero1
Gsta3	3.89871251249365e-44	0.805279610297088	0.889	0.528	3.11897000999492e-42	Entero1
Fabp2	1.29229051817192e-42	0.729193895321954	0.963	0.745	1.03383241453754e-40	Entero1
Spink1	3.66097718899745e-42	0.629779702665184	0.87	0.454	2.92878175119796e-40	Entero1
Arg2	7.98376218566301e-40	0.831847381415513	0.815	0.455	6.38700974853041e-38	Entero1
Apoa4	1.02016690530616e-39	0.61007480062237	0.838	0.442	8.16133524244929e-38	Entero1
Mgst2	7.10101680893117e-39	0.851374862110365	0.912	0.709	5.68081344714493e-37	Entero1
S100g	2.96870148861901e-37	0.67331354650757	0.894	0.539	2.37496119089521e-35	Entero1
Gsta1	1.35306424273308e-32	0.472486977624749	0.995	0.796	1.08245139418646e-30	Entero1
Tkfc	3.50611905883385e-32	0.411753509964946	0.94	0.687	2.80489524706708e-30	Entero1
Cyp4f14	5.38663195657445e-31	0.519981900345935	0.852	0.55	4.30930556525956e-29	Entero1
Ces2a	2.32096694803394e-30	0.625871228915386	0.991	0.83	1.85677355842715e-28	Entero1
Ugt2b5	1.16317697151327e-29	0.65462290216724	0.759	0.446	9.30541577210619e-28	Entero1
Cideb	1.66522526047928e-29	0.668807931062889	0.838	0.55	1.33218020838342e-27	Entero1
Khk	2.3562753913959e-29	0.559103120929781	0.912	0.691	1.88502031311672e-27	Entero1
Gm3776	6.14199666813945e-29	0.494460207582878	0.889	0.593	4.91359733451156e-27	Entero1
Ces1d	1.38677645463885e-28	0.400811817763043	0.653	0.28	1.10942116371108e-26	Entero1
Ndrg1	2.24776560017418e-27	0.521412321460839	0.87	0.601	1.79821248013934e-25	Entero1
Cyp3a11	2.06247550022557e-26	0.527798351284679	0.736	0.384	1.64998040018045e-24	Entero1
Cyp2b10	2.15539125536087e-25	0.697335344923729	0.704	0.387	1.7243130042887e-23	Entero1
Fam213a	6.10640077092692e-25	0.301752190721072	0.861	0.608	4.88512061674153e-23	Entero1
Cyp2c66	1.35874414549653e-23	0.488752146117802	0.736	0.512	1.08699531639722e-21	Entero1
Gstm4	1.47764282293861e-23	0.653896742719388	0.676	0.399	1.18211425835089e-21	Entero1
Ephx1	3.22919737315709e-20	0.456673877333623	0.806	0.531	2.58335789852567e-18	Entero1
Chp2	1.89761043469546e-19	0.411199981170197	0.764	0.508	1.51808834775637e-17	Entero1
Guca2b	6.3835695365107e-19	0.432207103895791	0.694	0.385	5.10685562920856e-17	Entero1
Acaa1b	2.25648025784849e-18	0.495888699080675	0.634	0.364	1.80518420627879e-16	Entero1
Cyp2c29	5.3831327465173e-18	0.394391182028341	0.597	0.299	4.30650619721384e-16	Entero1
Leap2	9.71064455701969e-17	0.342055623616844	0.653	0.361	7.76851564561575e-15	Entero1
Cyp2d26	1.33095364656475e-16	0.338288348487044	0.667	0.379	1.0647629172518e-14	Entero1

Mogat2	1.97346003982513e-16	0.58554835926939	0.662	0.412	1.57876803186011e-14	Entero1
Apoc3	1.7352844842063e-15	0.271366484121131	0.616	0.309	1.38822758736504e-13	Entero1
Gsta2	2.62559594439782e-15	0.250857127832048	0.69	0.435	2.10047675551825e-13	Entero1
Golgb1	1.29123859560501e-13	0.438636271560022	0.833	0.63	1.03299087648401e-11	Entero1
Cyp3a13	1.30217888157126e-13	0.394118685929322	0.671	0.476	1.04174310525701e-11	Entero1
Mt4	2.18300364431703e-08	0.345660220432632	0.551	0.388	1.74640291545362e-06	Entero1
Anxa2	0.000105107960280338	0.376798010321461	0.762	0.761	0.00840863682242703	PIC1
Rn7sk	4.20752227853046e-22	0.49105710835518	0.919	0.651	3.36601782282437e-20	PIC2
Pla2g2a	3.19123918126354e-40	1.3249929078319	0.913	0.564	2.55299134501083e-38	Stem
1110028F11Rik	3.66737905051025e-40	1.04343585788738	0.846	0.46	2.9339032404082e-38	Stem
Lgals1	7.35974158937406e-40	1.2525285566728	0.885	0.483	5.88779327149925e-38	Stem
Bex1	4.07766717099372e-39	1.43857975339842	0.923	0.586	3.26213373679497e-37	Stem
Irx5	1.38160802990555e-30	1.01446376098122	0.788	0.529	1.10528642392444e-28	Stem
Igfbp4	5.45407120418318e-27	0.818091470946171	0.817	0.558	4.36325696334655e-25	Stem
Ier3	8.66238894755805e-21	0.902890680630629	0.817	0.621	6.92991115804644e-19	Stem
Areg	1.44349039107372e-16	0.995256095963472	0.894	0.709	1.15479231285898e-14	Stem
Nucb2	4.38795737047699e-10	0.541151354998672	0.731	0.583	3.51036589638159e-08	Stem
Apoa4	5.93932200196654e-45	1.6534454453666	1	0.491	4.75145760157324e-43	Entero2
Sis	1.10343694212606e-44	1.92728989561058	1	0.586	8.82749553700845e-43	Entero2
Apoa1	1.1201725751272e-44	2.15854194457433	1	0.624	8.96138060101757e-43	Entero2
Fam213a	1.30222226270094e-43	1.92083136662646	1	0.636	1.04177781016075e-41	Entero2
Tkfc	7.58132539855834e-42	1.80677470413554	0.987	0.724	6.06506031884668e-40	Entero2
Aldob	4.37622510619118e-41	2.39954242291639	1	0.817	3.50098008495295e-39	Entero2
Prap1	8.67990158043496e-41	1.85693038324859	1	0.791	6.94392126434797e-39	Entero2
Spink1	4.84903542738247e-40	1.48825704587288	0.974	0.512	3.87922834190598e-38	Entero2
Khk	3.00508823625212e-39	1.34034925225687	1	0.719	2.40407058900169e-37	Entero2
Fabp1	5.99265196134052e-39	2.12550630530773	1	0.716	4.79412156907241e-37	Entero2
Adh1	1.82859672267536e-38	1.59803619061465	1	0.695	1.46287737814029e-36	Entero2
Gsta2	2.72420461425157e-37	1.14631403188223	0.961	0.452	2.17936369140126e-35	Entero2
Ccl25	6.84495230843095e-37	1.27948413836236	1	0.739	5.47596184674476e-35	Entero2
Ephx1	1.52344584782225e-35	1.21448812092998	0.987	0.559	1.2187566782578e-33	Entero2
Adh6a	2.20786040162389e-34	1.12917276504955	0.987	0.476	1.76628832129911e-32	Entero2
Cyp4f14	2.08329178298781e-33	1.21375095416519	0.987	0.587	1.66663342639025e-31	Entero2
Gsta1	2.59753045339481e-33	1.49463348002007	1	0.827	2.07802436271585e-31	Entero2
S100g	3.10615120049419e-33	1.12431617027042	0.974	0.589	2.48492096039535e-31	Entero2
Cyp2c66	5.35653099425473e-33	1.15968719761029	0.961	0.528	4.28522479540378e-31	Entero2
Chp2	1.02267027061363e-32	1.14689618487639	0.974	0.53	8.18136216490906e-31	Entero2
Cyp2d26	1.41893843629184e-32	1.03781759437069	0.934	0.402	1.13515074903347e-30	Entero2
Ndrg1	9.50957333894437e-32	1.13147307348477	0.961	0.636	7.60765867115549e-30	Entero2
Leap2	9.79198061789397e-32	1.02395326084673	0.882	0.388	7.83358449431517e-30	Entero2
Cyp3a11	3.84843836479325e-31	1.24933472690356	0.895	0.427	3.0787506918346e-29	Entero2
Guca2b	3.07316875882886e-30	1.01512661269658	0.908	0.417	2.45853500706308e-28	Entero2
Gm3776	8.43962680052236e-30	1.29978848049862	0.987	0.632	6.75170144041789e-28	Entero2

Ugt2b5	1.58240974168786e-29	1.06991560499757	0.921	0.482	1.26592779335029e-27	Entero2
Reg3a	6.55267691058194e-29	1.35713088197254	0.829	0.369	5.24214152846556e-27	Entero2
Apoc3	7.76643242226385e-29	0.985880738820278	0.842	0.338	6.21314593781108e-27	Entero2
Cyp3a13	3.55897381144747e-28	1.01591555496027	0.908	0.487	2.84717904915798e-26	Entero2
Fabp2	3.37223145568505e-26	0.887488937761581	0.987	0.778	2.69778516454804e-24	Entero2
Cideb	5.18165254326076e-26	0.929781489565813	0.974	0.585	4.14532203460861e-24	Entero2
Aldh1a1	2.55441426054888e-25	0.893818129793693	0.961	0.7	2.0435314084391e-23	Entero2
Ces1f	5.35862461659871e-22	0.851231933945787	0.961	0.59	4.28689969327897e-20	Entero2
Gsta3	1.30110417219731e-17	0.63389982862055	0.947	0.581	1.04088333775785e-15	Entero2
Mogat2	5.54585767826874e-16	0.674414589926087	0.789	0.441	4.436686142615e-14	Entero2
Reg1	4.54564681142072e-15	0.888576028125212	0.803	0.555	3.63651744913658e-13	Entero2
Mir22hg	1.33018419475433e-14	0.366380428724304	0.711	0.413	1.06414735580346e-12	Entero2
Arg2	3.26751048825669e-13	0.594244489315401	0.816	0.513	2.61400839060536e-11	Entero2
Reg3b	5.40722600169337e-13	0.693131482994203	0.855	0.585	4.3257808013547e-11	Entero2
Cyp2b10	1.12509349975993e-12	0.688885456222288	0.75	0.434	9.00074799807947e-11	Entero2
Ces2a	7.06803557675554e-12	0.396131534280798	1	0.855	5.65442846140443e-10	Entero2
Reg3g	1.30407635700036e-10	0.31181532413503	0.882	0.636	1.04326108560029e-08	Entero2
Ephx2	1.70167582930346e-10	0.437834332901168	0.684	0.42	1.36134066344277e-08	Entero2
Mgst2	2.72676342101281e-09	0.49248922756913	0.934	0.74	2.18141073681025e-07	Entero2
Acaa1b	1.43762853426396e-08	0.386561090746002	0.684	0.403	1.15010282741117e-06	Entero2
Mt4	1.45515610140142e-07	0.271929123109072	0.632	0.407	1.16412488112114e-05	Entero2
Ly6a	2.01002501530632e-36	1.80821629971927	0.926	0.527	1.60802001224506e-34	RS
Anxa2	1.02615198797397e-22	1.04409160576179	0.971	0.745	8.20921590379172e-21	RS
Gm3776	3.18379144779195e-20	0.950402325556938	0.956	0.637	2.54703315823356e-18	RS
Anxa1	9.29314805494896e-19	1.29093095938257	0.765	0.521	7.43451844395917e-17	RS
Gsta1	7.71481200331918e-18	0.858095925736984	1	0.829	6.17184960265535e-16	RS
Areg	1.07592965296389e-06	0.374112922663416	0.838	0.721	8.60743722371114e-05	RS
Olfm4	7.34682088759632e-38	2.10864807562475	0.97	0.558	5.87745671007706e-36	Olfm4
Stra6l	3.32234482841291e-09	0.330277790744615	0.716	0.544	2.65787586273033e-07	Olfm4
Reg3b	6.19284589582374e-21	1.79285294751421	0.939	0.588	4.95427671665899e-19	Paneth
Reg3g	4.23085976422072e-20	1.61610736104673	0.918	0.642	3.38468781137658e-18	Paneth
Pla2g2a	5.53846330858548e-14	1.11076702995807	0.837	0.589	4.43077064686839e-12	Paneth
Lgals1	2.42086721115051e-12	0.625986886999084	0.816	0.511	1.9366937689204e-10	Paneth
Bex1	3.87275526447042e-12	0.781270009856096	0.878	0.609	3.09820421157633e-10	Paneth
Ier3	3.95525301117695e-10	0.577960555043002	0.816	0.633	3.16420240894156e-08	Paneth
Igfbp4	0.000137577038169854	0.459617953534454	0.653	0.583	0.0110061630535883	Paneth
Spink4	3.52440643770931e-28	2.97630334738867	1	0.62	2.81952515016745e-26	Goblet
Tff3	4.23028954766408e-27	2.52408433135767	1	0.844	3.38423163813126e-25	Goblet
Fcgbp	1.58233080983152e-26	2.15228915927589	0.956	0.412	1.26586464786522e-24	Goblet
Agr2	4.13763902264459e-22	2.22989599251811	0.956	0.592	3.31011121811567e-20	Goblet
A930004J17Rik	7.15243251983664e-16	0.270159357631491	0.8	0.444	5.72194601586931e-14	Goblet
Guca2a	3.66076196916109e-15	1.5382791422853	0.867	0.562	2.92860957532887e-13	Goblet
Nucb2	4.38484098082501e-15	0.84224961248207	0.889	0.584	3.50787278466001e-13	Goblet

Clps	1.46445135819238e-14	1.43210436908701	0.822	0.486	1.1715610865539e-12	Goblet

Supplementary Table 3: Buffer-dependent-dissociation efficiencies for intestinal crypts. Summary of the single intestinal crypt dissociation yield using distinct buffer compositions, as indicated. All dissociations were performed in PBS and in the presence of Dnase I. The number of single cells and multiplets (representing doublets, triplets or even cell clumps) were counted. The buffer composition used for the single crypt profiling experiments (**Figure 3**, **Supplementary Figures 4&5**) is highlighted in "green". On average, 3 – 44 cells per crypt were recovered out of an estimated 250 cells[26], thus yielding a dissociation efficiency up to 20%. PBL, Protease *B. licheniformis*; ACC, Accutase; ED, EDTA; EG, EGTA; Mul, multiplet; S, Sarkosyl; SD, standard deviation; Sig., Sigma Accutase Cat. #A6964; Sin, singlet; T, Trypsin; Ther. Thermofisher Accutase Cat. #A1110501, TLES; TrpyLE select.

Buffer composition:							
PBL	ACC	T	TLE	EG	ED	S	**mean±SD**
	Sig.						4 single crypts | Sin = 5±5 | Mul = 3±3
			+				4 single crypts | Sin = 6±2 | Mul = 7±3
+							4 single crypts | Sin = 5±4 | Mul = 4±1
+		**+**					4 single crypts | Sin = 6±7 | Mul = 9±2
	Sig.			**+**	**+**		4 single crypts | Sin = 1±1 | Mul = 2±1
		+		**+**	**+**		4 single crypts | Sin = 4±1 | Mul = 3±3
+	**Sig.**			**+**	**+**		4 single crypts | Sin = 4±3 | Mul = 2±2
+	**Ther.**			**+**	**+**		21 single crypts | Sin = 16±6 | Mul = 7±3
		+		**+**	**+**		4 single crypts | Sin = 7±4 | Mul = 3±3
+		**+**		**+**	**+**	**+**	4 single crypts | Sin = 5±2 | Mul = 3±3