Deterministic scRNA-seq of individual intestinal organoids reveals new subtypes and coexisting distinct stem cell pools

Johannes Bues[1,2,8], Marjan Biočanin[1,2,8], Joern Pezoldt[1,2], Riccardo Dainese[1,2], Antonius Chrisnandy[3], Saba Rezakhani[3], Wouter Saelens[4,5], Revant Gupta[2,6], 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 and Marjan Biočanin

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

Abstract

While Single-cell RNA-sequencing (scRNA-seq) has transformed our ability to resolve cellular properties across systems, current microfluidics-based scRNA-seq technologies are limited to samples with large amounts of cells (> 1,000 cells). This prevents efficient processing of individual, small tissue samples, leading to confounded mosaic cell population read-outs. To overcome these limitations, we developed a deterministic, mRNA-capture bead and cell co-encapsulation dropleting system, DisCo, that enables precise particle and cell positioning and droplet sorting control through combined machine-vision and multilayer microfluidics. To underscore the unique capabilities of our approach in processing low-input samples (< 100 cells) at high efficiency (> 70%), we analyzed intestinal organoid development by “DisCo-ing” 31 individual organoids at varying developmental stages. This revealed extensive organoid heterogeneity, uncovering a so far uncharacterized “gobloid” subtype consisting predominantly of precursor and mature (*Muc*2^{+}) goblet cells. We also identified spheroids that are comprised of a secondary regenerative fetal-like *Ly6a*$^{+}$ stem cell population, marked by prolonged YAP1 nuclear translocation and target gene expression, and that persist as symmetrical cysts even under differentiation conditions. These findings illustrate the power of our “no-cell-left-behind” platform in providing high-resolution snapshots of cellular heterogeneity among individual, small 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[2–5], have led to a substantial increase in experimental throughput, enabling massive mapping efforts such as the mouse and human cell-atlas studies[5–7]. These developments were accompanied by biochemical advances, for instance for targeted transcript detection or library multiplexing[8,9], and together 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. Consequently, small samples, involving for instance zebrafish embryos[10], organisms like *C. elegans*[11], or intestinal organoids[12,13], tend to still be pooled to reach cell numbers that are compatible with stochastic 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 developed a deterministic droplet-based scRNA-seq approach (DisCo), eliminating current cell input limitations. In contrast to other stochastic cell capture methods, we implemented an active particle coordination approach that allows for precise capture and co-encapsulation of one cell and one mRNA-capture bead. We demonstrate that DisCo can efficiently process samples of 100 cells and below, making this “no-cell-left-behind” approach well suited to handle individual, small tissues. Here, we exploit DisCo’s unique capabilities to explore the heterogeneous development of single 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[12]. These unpredictable developmental patterns represent one of the major limitations of this model system, preventing widespread implementation e.g. in drug screens[14]. 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[15,16], are of essence. In depth mapping of individual organoid heterogeneity by scRNA-seq has been so far 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[12]. In total, we “DisCo’d” 31 single organoids at 4 developmental time points post symmetry breaking, and identified striking differences in cell type composition between organoids. Among the identified subtypes, we uncovered a rare phenotype predominantly comprised of precursor- and mature goblet cells, which we term “gobloids”. In addition, we detected “spheroids” that are composed of regenerative fetal-like stem cells marked by Stem Cell Antigen-1 (Sca-1/Ly6a)[17–20] and that persist even under differentiation conditions, likely driven by YAP1-driven mechanosensing[21].

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 valve[22], 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 exercise, 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 in droplet scRNA-seq co-encapsulation purity and cell capture efficiency are critical system parameters, 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 around 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 a previously developed chip-based cDNA generation protocol[23]. Initially, as a library quality measure, we performed a species-mixing experiment of human HEK 293T and murine brown pre-adipocyte IBA cells, and found that this generates scRNA-seq data with clear species separation (**Figure 1H**), consistent with the limited number of previously detected doublets (**Figure 1F**). However, the number of cell barcodes that we detected vastly exceeded the number of cells during the experiment (approximately 200 cells). To investigate this, we revisited the scRNA-seq data, revealing a skewed barcode sequence hamming distance distribution compared to a true random distribution (**Supplementary Figure 1F**). Thus, our data suggested a statistically unlikely presence of barcodes with only 1-2 nt differences. Since the uniquely low number of beads in DisCo samples (< 500) renders the random occurrence of barcode sequences with a hamming distance < 3 rare, we hypothesized that highly similar barcodes could be unified. To test this, we merged barcodes of the species mixing data at various hamming distance thresholds, and found that such merging at a value of up to two did not compromise single cell purity (**Supplementary Figure 1G** and **1H**).

We observed that this barcode merger, besides reducing the number of barcodes to a relevant level, also improved the detectable number of transcripts per cell as compared to published Drop-seq datasets on HEK 293T cells[2,23], which were down-sampled to normalize the sequencing depth across different experiments (**Figure 1I**). As expected, due to the low number of failed co-encapsulation events, and thus to the low number beads that were not co-encapsulated with a cell, the DisCo system showed an improved read utilization rate (**Supplementary Figure 1I**).

Since DisCo is a stopped-flow system, we observed that the system requires negligible run-in time, and is capable of efficiently processing cells from the first cell on, hence the “no-cell-left-behind” label. 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 1J**). Utilizing the Dispencell approach, we processed cell numbers ranging between 50 - 200 cells, of which on average 86.6% (SD ± 8.1%) were visible on the chip. Of all input cells, 79.1% (SD ± 7.4%) were successfully co-encapsulated, which corresponds to a co-encapsulation efficiency of 91.6% (SD ± 1.6%), while 84.2% (SD ± 7.9%) of input cells were found as barcodes over 500 UMIs per cell (**Figure 1J**). Despite correcting for cell barcode errors and prefiltering of the data, we observed a slight overestimation of cell numbers (captured vs. sequenced cells: 79.1% SD± 7.4% vs. 84.2% SD ± 7.9%) most likely due to barcodes exceeding the hamming distance thresholds.

As a first real-world application, we used DisCo to explore the developmental heterogeneity of intestinal organoids[24]. These polarized epithelial tissues are generated by intestinal stem cells in 3D matrices through a stochastic self-organization process, and mimic key geometric,

architectural and cellular hallmarks of the adult intestinal mucosa (e.g. a striking crypt-villus-like axis). When grown from single stem cells, organoids of very different morphologies form under seemingly identical *in vitro* conditions (**Figure 2A,** overview image in **Supplementary Figure 2A**). Pooled tissue scRNA-seq efforts have shed light on the *in vivo*-like cell type composition of these organoids[12,13,25], but due to the limitation of cell numbers have not been able to 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 responsible for crypt formation[12]. Here, we were particularly interested in examining the emergence of heterogeneity between individual organoids subsequent to the symmetry breaking timepoints. To do so, we isolated single LGR5^{+} cells by FACS, and maintained them in a stem cell state using CHIR99021 and valproic acid (CV)[26]. 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, 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 cells detected on chip was noted and correlated to the number of barcodes retrieved, which was in approximate accordance (**Supplementary Figure 2B**). The observed even distribution of the number of ribosomal reads and low heat shock protein expression indicated 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 1132 sampled cells through Uniform Manifold Approximation and Projection (UMAP). We found that our data was consistent with previously published pooled organoid scRNA-seq read-outs[13,25] 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**).

Noteworthy, we found that batch effects are correctable since no batch-based clustering was observed after correction (**Supplementary Figure 2D**). We also did not detect any clustering driven by cell quality, e.g. detected transcripts or mitochondrial transcripts (**Supplementary Figure 2E**). These findings support the cell type-resolving power of our DisCo platform (**Figure 2C,** extensive heatmap in **Supplementary Figure 2F,** and list in **Supplementary Table 1**). In addition, we observed a distinct cluster marked by high expression of Stem cell antigen 1 (*Sca1* or *Ly6a*), next to two remaining clusters that did not show a striking marker gene signature. To resolve the identity of the latter clusters, we imposed temporal information on the data. This revealed that these clusters likely represent stem- and previously termed potentially intermediate cells (PIC)[27], given their occurrence at early developmental time points (**Figure 2D**). As expected, mature cell types were mostly present at later time points. In contrast, *Ly6a*-expressing cells were identified at S0, S1, and S3, and high expression of *Anxa1* in the same cluster (**Supplementary Figure 2F**) suggested that these cells are regenerative fetal-like stem cells[18,19,28]. To further leverage the temporal component in the DisCo data, we used slingshot trajectory analysis[29] to infer lineage relationships between cell types. The reconstructed differentiation trajectories followed expected patterns with S0 stem cells directly differentiating into Paneth cells during symmetry breaking, and enterocytes and goblet cells differentiating from a common progenitor pool (**Figure 2E**). Finally, we utilized the lineage reconstruction to identify genes that may be of particular significance for each waypoint. Beyond the previously utilized marker genes for cell type annotation, for example *Reg3b* and *Reg3g* for Paneth cells, additional established markers[30] were identified, such as *Agr2*, *Fcgbp*, and *Clca1* for goblet cells (**Figure 2F**). Overall, this suggests that the meta-data produced with our DisCo platform aligns with prior knowledge.

To subsequently provide a unique snapshot of inter-organoid heterogeneity, we next stratified our cells according to the individual organoids from which they were derived by mapping this information onto the reference scaffold (**Figure 3A**). This analysis revealed that the maturation

and cell type composition of individual organoids seems to follow the expected pattern with early organoids (S0) mainly containing stem cells, as well as Paneth and *Ly6a*$^{+}$ cells, and older organoids (S1 – S3) differentiated cells like goblet cells and enterocytes. Nevertheless, we identified multiple organoids that were predominantly made up of single cell types. For example, two organoids were mainly comprised of *Ly6a*$^{+}$ cells (S1a, S3e), three contained mostly enterocytes (S2c, S3b, S3d), and two consisted predominantly of immature and mature goblet cells (S1b, S2f). A summary of cell type distribution per organoid, and marker genes is presented in **Supplementary Figure 3A** and **3B**, respectively. Furthermore, the same organoids showed aberrant morphologies compared to canonical organoids bearing a crypt-villus axis (e.g. S3c). Specifically, those containing mostly enterocytes (S3d) or goblet cells (S2f) tended to be small and round, while *Ly6a*$^{+}$ cell-containing organoids (S3e) exhibited a larger, cystic like structure (**Supplementary Figure 3C**). *Ly6a*$^{+}$ cell- and enterocyte-enriched organoids have been previously described as spheroids[18] and enterocysts[12], respectively. However, to our knowledge, organoids displaying goblet cell hyperplasia, here termed “gobloids”, were so far unknown. To further investigate inter-organoid heterogeneity, we visualized cells by marker gene expression for each individual organoid. Marker gene expression was consistent with previously identified organoid subtypes (Spheroid: *Ly6a*, Goblet cells: *Muc2*, Enterocysts: *Fabp1*) (**Figure 3B**). We subsequently used psupertime[31] to identify genes that are dynamically expressed during the development of individual organoids which revealed additional genes 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**). To validate the existence of the uncovered organoid subtypes, we utilized RNAScope to visualize *Ly6a*, *Muc2*, *Fabp1* expression in organoid sections (**Figure 3C,** controls in **Supplementary Figure 3E**). We found canonical budding organoids, containing few goblet cells and enterocytes, and no *Ly6a*$^{+}$ cells. *Ly6a*$^{+}$ cells were found in spherical organoids that did not contain cells expressing *Muc2* and *Fabp1*, which is why we defined these as spheroids. Similarly, we detected organoids

exclusively containing *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" (**Figure 3C**), which, based on our trajectory analyses (**Figure 2E**), may be derived from so far uncharacterized PICs.

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. To investigate this hypothesis, we performed flow cytometry and found indeed that the majority of cells are either LGR5^{+} LY6A^{-} (24.5 %) or LGR5^{-} LY6A^{+} (3.3 %) with only a minority (0.44%) being double positive (**Figure 4A**). We next aimed to determine if LGR5^{-} LY6A^{+} cells were able to generate organoids. We found that both LGR5^{+} LY6A^{-} and LGR5^{-} LY6A^{+} cells produce organoids of similar morphological heterogeneity (**Figure 4B**), indicating full stem cell potential that is comparable to that of previously described fetal-like stem cells[18]. Furthermore, as LGR5^{-} LY6A^{+} cells did not display a propensity towards spheroid formation, we hypothesized that LY6A^{+} cells revert into a state that is capable of producing canonical organoids. In order to explore this hypothesis, we leveraged lineage reconstruction to identify differentiation trajectories for *Ly6a*$^{+}$ cell differentiation. This analysis indicated that *Ly6a*$^{+}$ cells indirectly differentiate by passing through the canonical stem cell state (**Figure 4C**, arrows).

Given that both stem cell populations form comparably heterogeneous organoid cultures, we speculated that spheroid formation is not resulting from an innate property of the seeded cells, but rather is triggered by an environmental stimulus. Previous research suggested that mechanosensing mediated by YAP1 can impact the capacity for symmetry-breaking of organoids[12,21]. Thus, we next analyzed YAP1 gene regulatory activity across organoids in our scRNA-seq data. These analyses revealed low YAP1 regulatory activity in differentiated cells, given that genes that were reported to be repressed by YAP1[28] showed higher expression,

especially in the enterocyte cluster (**Supplementary Figure 4**). In contrast, we observed highest YAP1 regulatory activity in *Ly6a*-expressing spheroids given that the expression of genes that are known to be activated by YAP1[12] is most prominently increased in this type of organoid (**Figure 4C**). These findings support the involvement of this transcriptional co-regulator in spheroid formation. To further investigate the potential impact of mechanosensing, we seeded single cells on 2D layers of Matrigel for 24 hours (**Figure 4D**)[32,33], a configuration that induces YAP1 nuclear localization[34]. After 24 hours of induction, we added Matrigel on-top to implement 3D culture conditions. Interestingly, this treatment induced the formation of thin epithelial spheroids at a higher rate than observed in a conventional culture (**Figure 4E**). We further tested the molecular identity of these cells by flow cytometry, and found a significantly greater number of LY6A^{+} cells in these cultures compared to standard conditions (**Figure 4F**). This observation suggests that these cells are indeed resembling *Ly6a*$^{+}$ cells, as observed in spheroids.

As our data invoked YAP1-mediated gene activation is involved in spheroid formation, we further analyzed these cultures by localizing the spatial distribution of YAP1 in canonical organoids and spheroids by immunocytochemistry. Indeed, we found YAP1 to be cytoplasmic in canonical organoids, and predominantly localized to the nucleus in thin epithelial spheroids (**Figure 4G,** S0). Spheroids morphologies were identified in non-differentiation conditions (S0), as well as in differentiation conditions (S3), both showing more nuclear YAP1 localization compared to conventional culture (**Figure 4G,** S3). Noteworthy, only a minority of organoids retained the spheroid morphology under differentiation conditions. We next explored the differentiation state of sandwich-cultured organoids. Under non-differentiation (S0) conditions, we were able to identify both spheroid and canonical phenotypes, marked by LY6A high and low expression, respectively. As expected, the absence of ALDOB expression indicated the lack of enterocytes in both organoid types under expansion conditions (**Figure 4H,** S0). Under differentiation conditions (S3), we found that LY6A-low canonical organoids were budding and producing ALDOB-

expressing enterocytes, while spheroids with visibly higher LY6A showed no enterocyte formation (**Figure 4H,** S3). Overall, our experimental results indicate that LY6A-high stem cell spheroid formation is marked by higher YAP1 nuclear localization. Under differentiation conditions, spheroids have the capacity to retain their spheroid state.

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 "no-cell-left-behind" approach provides the advantage of being able to process low cell input samples, thus filling an important gap in the scRNA-seq toolbox. Moreover, the stopped-flow nature of our platform is offsetting to some extent the cost of its decreased throughput, as it allows for continuous operation of the system. 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[35,36]. In concept, DisCo is thus fundamentally different to passive particle pairing approaches such as traps[37–39] and, compared to these technologies, offers the advantage of requiring vastly simpler chips without suffering from cell/particle size and shape selection bias[40,41]. This renders the DisCo approach universally applicable to any particle co-encapsulation application[42,43], 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 cell numbers.

To demonstrate DisCo's capacity to process small tissues/systems that were so far difficult to access experimentally, we have so far applied it to analyze the cell heterogeneity of

chemosensory organs from *Drosophila* larvae (Maier et al., BioRxiv, 2020) and, as shown here, single intestinal organoids. The latter led us to uncover organoid subtypes of aberrant cell type distribution that were previously not resolved with pooled organoid scRNA-seq[12,13,25]. Of particular interest is one organoid subtype that we termed “gobloid" given that it predominantly comprises immature and mature goblet cells. Since low Notch signaling is pivotal for lineage commitment of crypt base columnar (CBC) cells towards secreting progenitors, lack of Notch-ligand providing Paneth cells[44], may drive gobloid development[45]. However, failure to produce Paneth cells has previously been suggested as the mechanism for enterocyst development[12], 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. We also identified a secondary stem cell population marked by *Ly6a* expression, nuclear localized YAP1, and high YAP1 target activity. These cells were previously described as fetal-like stem cells or revival stem cells occurring during intestinal regeneration[19,20,28]. By mapping the *Ly6a*-linked cell state to individual organoids in our DisCo data, we found the respective cells to be vastly overrepresented in a subset of organoids, which, both morphologically and molecularly, show a striking similarity to previously described spheroid-type organoids[17,18]. Hence, we hypothesize that beyond its role in symmetry breaking[12], high YAP1 signaling reprograms stem cells to *Ly6a+* stem cells during the same period[28]. Indeed, organoids composed of only *Ly6a+* stem cells do not break symmetry and continue their development as spheroids. Furthermore, we found that the initial cell state during seeding is not indicative for spheroid formation. Rather, a strong prolonged mechanical stimulus appears required to trigger cellular reprogramming[21]. This suggests that YAP1 mechanosensing is induced by variations in matrix properties[15], which leads to differences in development.

In sum, we demonstrate that scRNA-seq of individual organoids is a powerful approach to explore tissue heterogeneity and to yield mechanistic insights into how this heterogeneity arises. 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 with the engineering of more robust organoid systems. Furthermore, we expect this approach to be applicable to other developing systems or small, clinical samples to gain detailed insights into developmental robustness and cellular heterogeneity.

Acknowledgements

We thank Wanze Chen, Petra C. Schwallie, and Vincent Gardeux 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 an Animalfree Research 3R Grant, 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 chip. JB developed machine-vision integration for the DisCo. JB and MB benchmarked the system and performed all single-cell RNA-seq experiments. JP, JB, MB, WS and RG performed data analysis related to single organoid scRNA-seq experiments. JB, JP, and MB performed the sandwich culture and flow cytometry assays. JB, SR and MB performed all organoid and cell culture assays. JB, AC, and JR, performed all imaging assays. 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. 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).
5. Han, X. *et al.* Mapping the Mouse Cell Atlas by Microwell-Seq. *Cell* **172**, 1091–1107 (2018).
6. The Tabula Muris Consortium. Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. *Nature* **562**, 367–372 (2018).
7. 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
8. Saikia, M. *et al.* Simultaneous multiplexed amplicon sequencing and transcriptome profiling in single cells. *Nat. Methods* **16**, 59–62 (2019).
9. 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).
10. Wagner, D. E. *et al.* Single-cell mapping of gene expression landscapes and lineage in the zebrafish embryo. *Science* **360**, 981–987 (2018).
11. Packer, J. S. *et al.* A lineage-resolved molecular atlas of C. Elegans embryogenesis at single-cell resolution. *Science* **365**, (2019).
12. Serra, D. *et al.* Self-organization and symmetry breaking in intestinal organoid development. *Nature* **562**, 66–72 (2019).
13. Grün, D. *et al.* Single-cell messenger RNA sequencing reveals rare intestinal cell types. *Nature* **525**, 251–255 (2015).
14. Rossi, G., Manfrin, A. & Lutolf, M. P. Progress and potential in organoid research. *Nat. Rev. Genet.* **19**, 671–687 (2018).
15. Gjorevski, N. *et al.* Designer matrices for intestinal stem cell and organoid culture. *Nature* **539**, 560–564 (2016).
16. Brassard, J. A. & Lutolf, M. P. Engineering Stem Cell Self-organization to Build Better Organoids. *Stem Cell* **24**, 860–876 (2019).
17. Mustata, R. C. *et al.* Identification of Lgr5-Independent Spheroid-Generating Progenitors of the Mouse Fetal Intestinal Epithelium. *Cell Rep.* **5**, 421–432 (2013).
18. Yui, S. *et al.* YAP / TAZ-Dependent Reprogramming of Colonic Epithelium Links ECM Remodeling to Tissue Regeneration. *Cell Stem Cell* **22**, 35–49 (2018).
19. Ayyaz, A. *et al.* Single-cell transcriptomes of the regenerating intestine reveal a revival stem cell. *Nature* **569**, 121–125 (2019).
20. Roulis, M. *et al.* Paracrine orchestration of intestinal tumorigenesis by a mesenchymal niche. *Nature* **580**, 524–529 (2020).
21. Dupont, S. *et al.* Role of YAP/TAZ in mechanotransduction. *Nature* **474**, 179–184 (2011).
22. 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).
23. 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).
24. Sato, T. *et al.* Single Lgr5 stem cells build crypt – villus structures in vitro without a mesenchymal niche. *Nature* **459**, 262–266 (2009).
25. Haber, A. L. *et al.* A single-cell survey of the small intestinal epithelium. *Nature* **551**, 333–339 (2017).
26. Yin, X. *et al.* Niche-independent high-purity cultures of Lgr5 + intestinal stem cells and their progeny. *Nat. Methods* **11**, 106–112 (2014).
27. Battich, N. *et al.* Sequencing metabolically labeled transcripts in single cells reveals mRNA turnover strategies. *Science* **367**, 1151–1156 (2020).
28. 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).
29. Street, K. *et al.* Slingshot : cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics* **19**, 1–16 (2018).
30. 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).
31. Macnair, W. & Claassen, M. psupertime: supervised pseudotime inference for single cell RNA-seq data with sequential labels. *bioRxiv* 622001 (2019). doi:10.1101/622001
32. Tong, Z. *et al.* Towards a defined ECM and small molecule based monolayer culture system for the expansion of mouse and human intestinal stem cells. *Biomaterials* **154**, 60–73 (2018).
33. Altay, G., Larrañaga, E., Tosi, S., Barriga, F. M. & Batlle, E. Self-organized intestinal epithelial monolayers in crypt and villus-like domains show effective barrier function. *Sci. Rep.* 1–14 (2019). doi:10.1038/s41598-019-46497-x
34. Aragona, M. *et al.* A mechanical checkpoint controls multicellular growth through YAP/TAZ regulation by actin-processing factors. *Cell* **154**, 1047–1059 (2013).
35. 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).
36. 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).
37. 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).
38. 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).
39. Zhang, M. *et al.* Highly parallel and efficient single cell mRNA sequencing with paired picoliter chambers. *Nat. Commun.*

11, 2118 (2020).
40. DeLaughter, D. M. The Use of the Fluidigm C1 for RNA Expression Analyses of Single Cells. *Curr. Protoc. Mol. Biol.* **122**, 1–17 (2018).
41. 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).
42. Dura, B. *et al.* Profiling lymphocyte interactions at the single-cell level by microfluidic cell pairing. *Nat. Commun.* **6**, (2015).
43. Gérard, A. *et al.* High-throughput single-cell activity-based screening and sequencing of antibodies using droplet microfluidics. *Nat. Biotechnol.* (2020). doi:10.1038/s41587-020-0466-7
44. Sato, T. *et al.* Paneth cells constitute the niche for Lgr5 stem cells in intestinal crypts. *Nature* **469**, 415–418 (2011).
45. 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).

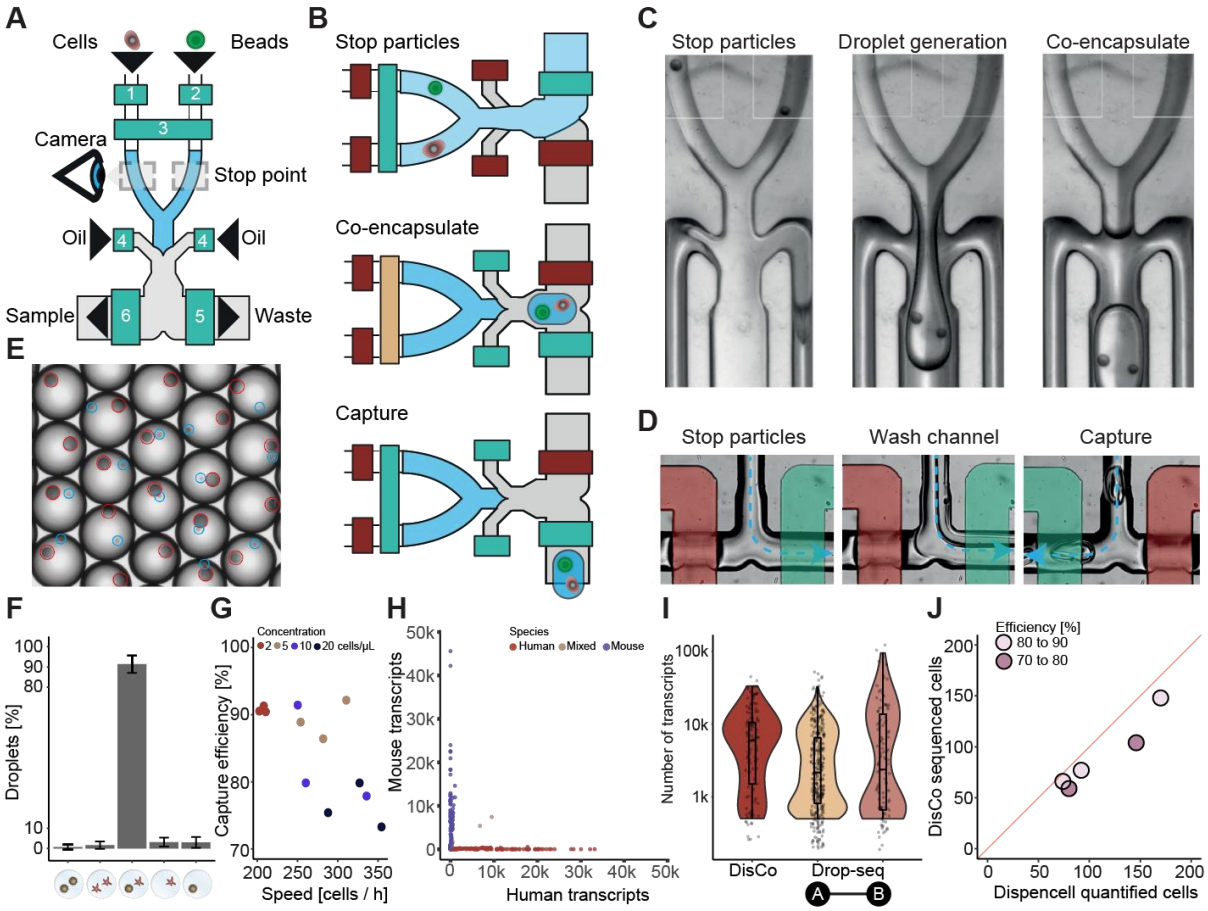

Figure 1 Overview and critical feature assessment of the deterministic co-encapsulation (DisCo) system: (**A**) Schematics of the DisCo microfluidic device. The device contains three inlet channels for cells, beads, and oil, and two outlets for waste and sample liquids. All inlets and outlets are augmented with Quake valves (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 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, 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 washed prior to droplet capture. Upon co-encapsulation, the waste valve is closed and 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. Error bars represent standard deviation. (**G**) Cell capture efficiency and cell capture speed for varying cell concentrations. 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 iBA cells were processed with the DisCo workflow for scRNA-seq, and species separation visualized as a Barnyard plot. Barcodes were merged at hamming distance 2. (**I**) Comparison of detected UMIs per cell of conventional Drop-seq experiments. UMIs per cell from HEK 293T data for conventional Drop-seq experiments (A - from Biočanin, Bues *et al.* 2019[23] and B - 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. (**J**) Total cell processing efficiency of DisCo at low cell inputs. Different numbers of input, ranging from 74 to 170 sorted HEK 293T cells, were quantified with the Dispencell system. Subsequently, all cells were processed with DisCo, sequenced, and quality filtered. The red line represents 100% efficiency, and samples were colored according to recovery efficiency after sequencing.

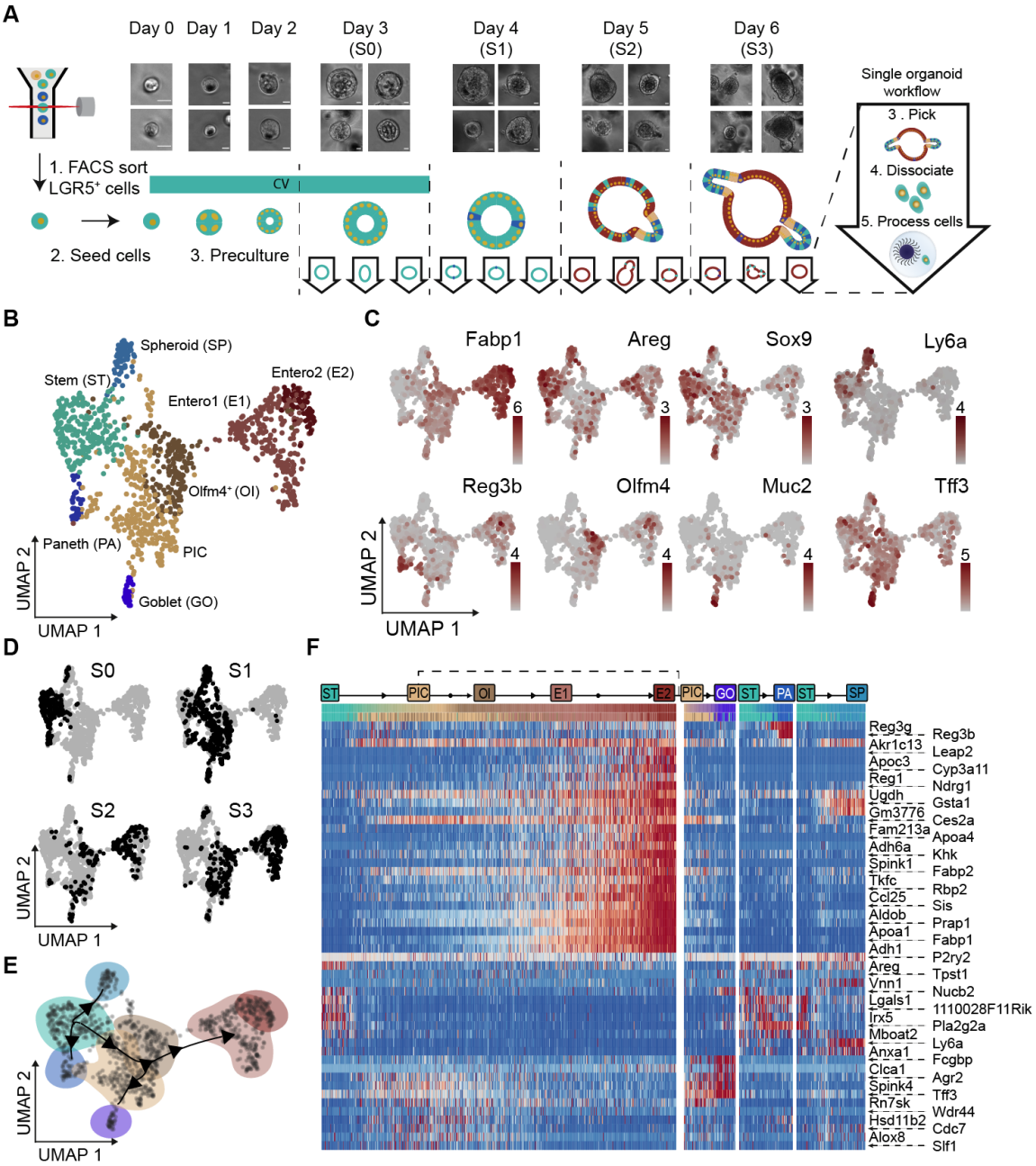

Figure 2 Utilizing DisCo to reconstruct single intestinal organoid heterogeneity along development: (**A**) Overview of the experimental design for DisCo'ing individual organoids. Single LGR5^{+} intestinal stem cells were isolated via FACS and precultured for 3 days under stem cell maintenance conditions (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 50 µm. (**B**) UMAP embedding of all sequenced cells. All 1132 processed cells from 31 organoids were clustered with k-means clustering, after which clusters were annotated according to marker gene expression. (**C**) Gene expression on UMAP embedding 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[29]. 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), Potential intermediate cells (PIC)[27], Enterocytes cluster 1/2 (Entero1/2).

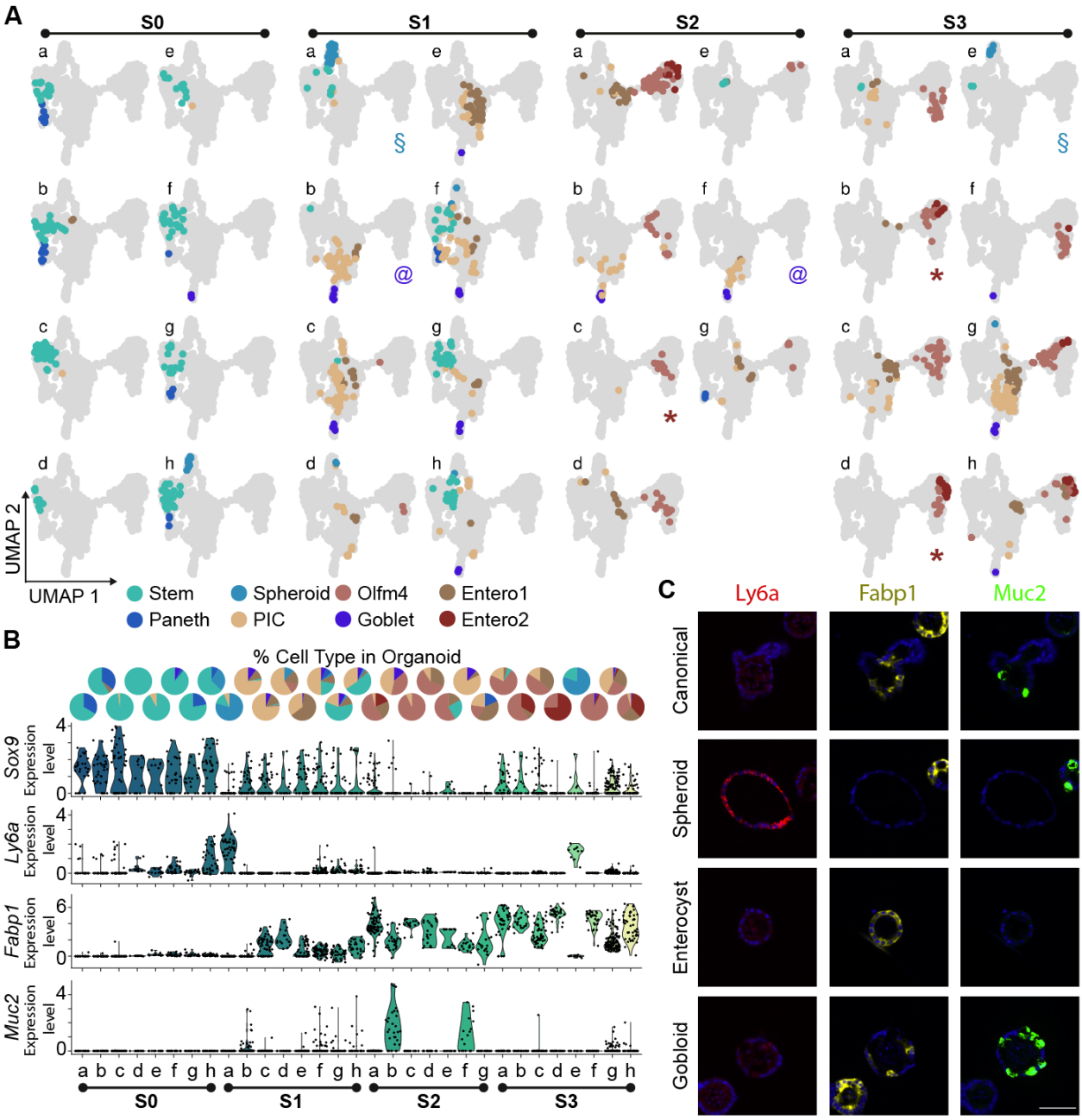

Figure 3 Cell type distribution and marker gene expression across individual intestinal organoids during development: (**A**) Projection of cell types onto 31 individual organoids. Cells per single organoid were colored according to their global clustering and highlighted on the UMAP embedding of all sequenced cells. Projections are grouped according to their sampling time. Manually classified organoids were annotated with the following symbols: "*" enterocysts, "§" spheroids, "@" gobloids. (**B**) Expression of key marker genes (scaled total UMI counts per cell) across individual organoids grouped according to their sampling time. Two stemness markers (*Sox9* and *Ly6a*), one enterocyte (*Fabp1*) and one goblet cell (*Muc2*) are shown. Pie charts summarize the cell type abundance per individual organoid. (**C**) *in situ* RNA detection for identified organoid-specific markers. RNA fluorescence in situ hybridization (RNAscope) of organoid subtypes showing RNA expression of *Ly6a, Fabp1,* and *Muc2* in representative images for spheroids, enterocysts, gobloids, and canonical intestinal organoids. Scale bar 50 µm.

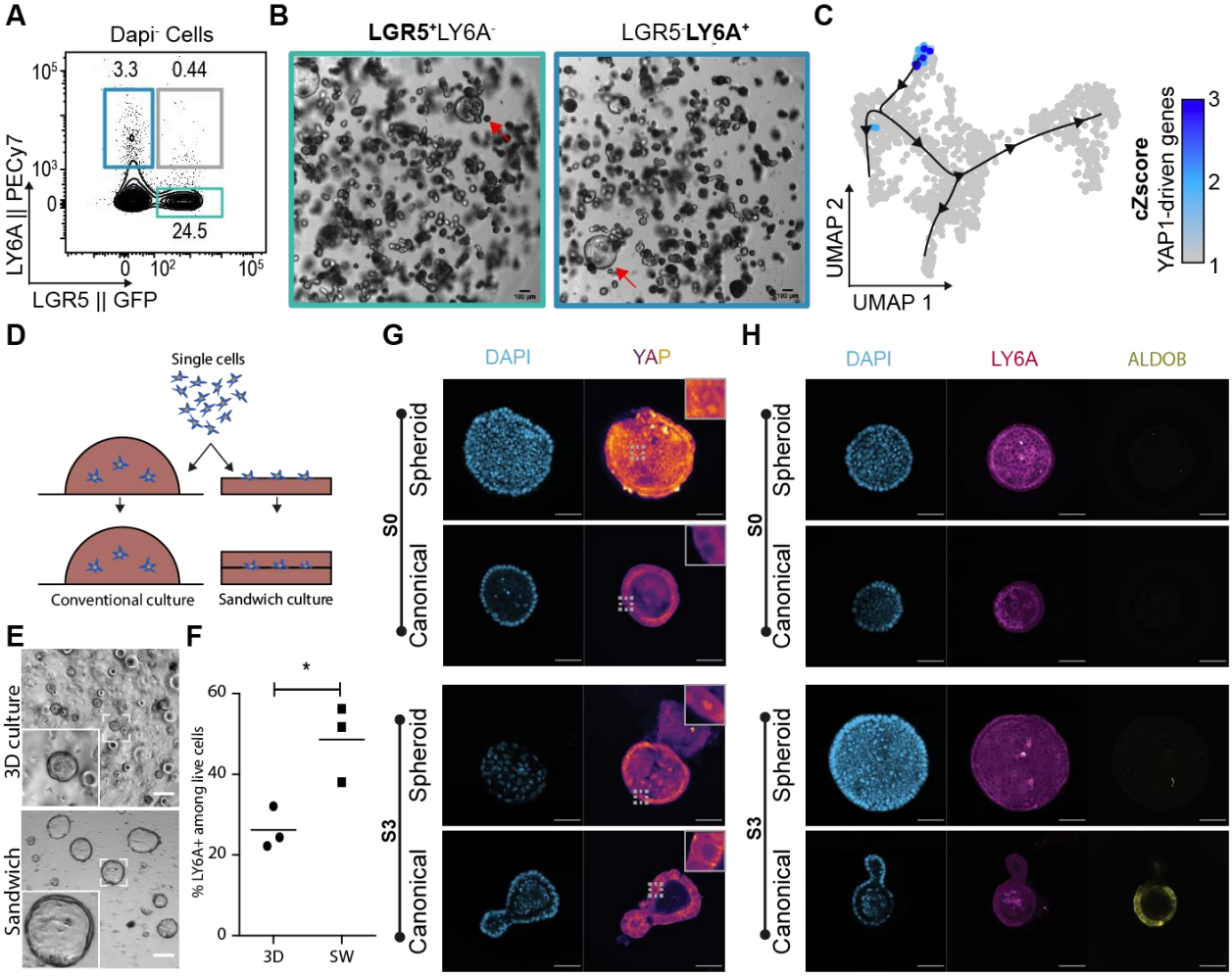

Figure 4 Role of YAP1 in Ly6a-high spheroids and canonical organoids in stem-cell and differentiation conditions: (**A**) Surface LY6A and LGR5-GFP expression under ENR CV conditions. The dot plot depicts LGR5-GFP and Ly6a expression in organoid-derived single cell suspensions. The numbers indicate frequencies in percent. (**B**) 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. (**C**) UMAP embedding with slingshot trajectories[29] calculated from the *Ly6a*$^+$ cluster as starting point, and YAP1 target gene activity onto the UMAP embedding. The expression of genes that are positively regulated by YAP1[12] was calculated as the cumulative Z-score and projected on the UMAP embedding of all sequenced cells. (**D**) Illustration of a sandwich culture experiment. For the sandwich culture, single cells from intestinal organoids are seeded on top of a pre-casted Matrigel layer, and after 24 hours sealed with an additional layer of Matrigel. (**E**) Culturing outcome of sandwich and conventional cultures (3D). Precultured cells from common cultures were dissociated and cultured in sandwich and conventional cultures. Images were taken 48 hours post-seeding. Scale bar 100 µm. (**F**) LY6A expression in conventional (3D) and sandwich cultures (SW). Single cell suspensions were obtained from organoids 48 hours post-seeding and analyzed using flow cytometry. The scatterplot summarizes frequencies of LY6A$^+$ cells per condition. Significance was tested with a Mann–Whitney–Wilcoxon test ($P < 0.05$). (**G**) YAP1 localization in sandwich culture organoids. Sandwich cultures were prepared after which organoids were immuno-fluorescently labelled for YAP1 at S0 and S3. Images depict representative spheroid and canonical organoids. Scale bar 50 µm. (**H**) LY6A and ALDOB staining in sandwich culture organoids. As in panel G, sandwich cultures were prepared after which organoids were immuno-fluorescently stained for LY6A (violet) and ALDOB (yellow) at S0 and S3. Images depict representative spheroid and canonical organoids. Scale bar 50 µm.