Thèse n° 7887

Developing new tools to study and program cell fate at the single-cell level

Présentée le 15 juin 2021

Faculté des sciences de la vie
Unité du Prof. Deplancke
Programme doctoral en biotechnologie et génie biologique

pour l'obtention du grade de Docteur ès Sciences

par

Marjan BIOCANIN

Acceptée sur proposition du jury

Prof. D. Trono, président du jury
Prof. B. Deplancke, directeur de thèse
Dr P. Carninci, rapporteur
Prof. S. Aerts, rapporteur
Prof. D. Suter, rapporteur

École
polytechnique
fédérale
de Lausanne

2021

The passenger
(translation by Milica Špadijer)

I walk freely
Nobody took away from me
The freedom to kiss the sad eve.
I spread my arms not towards sunrise
But towards night and the sea.

I enter with a smile wherever I may come
Into realities painful and sad.
The sins, too, when I love
Weave the skies
Around my joyfully bent head.

I leave the dream with a painful smile
To pass, go and die
Love is an endless path
on which everything is allowed.

I don't feel sorry for either you or me
And I smile towards the distance
Only in my eye fatigue gleams
And all I ask from you
Is a quiet instance
Instance.

Miloš Crnjanski

Путник

Идем слободно
Нико ми није однео
Да љубим тужну ноћ.
Раширим руке, али не у зоре
него у море и ноћ.

Осмехом улазим, стигао ма куд,
У тужне и болне јаве.
Кад волим мени и греси свуд
Небеса плету,
Око радосно погнуте главе.

Остављам болним осмехом сан,
Да прође и оде и мре.
Љубав је пут бескрајан
На ком је дозвољено све.

Не жалим ни тебе ни себе ја,
И смешим се на даљине.
Умор ми само у очима сја,
И све што иштем од тебе
То је часак два тишине,
Тишине.

Милош Црњански

Acknowledgments

One of the most significant successes of my thesis was managing to surround myself with inspiring people from whom I could learn and who were my inspiration to endure, question, and grow further. These are the people to whom I am the most thankful.

First and foremost, I would like to thank Prof. Dr. Bart Deplancke for being a great mentor, for his patience, and supporting all that we built together from the beginning to the end of this thesis. I am thankful for his enthusiasm and daring me to do the unimaginable. You created a fantastic lab atmosphere, believed in our potential, and shaped our growth.

I want to thank Prof. Dr. Didier Trono, Prof. Dr. David Suter, Prof. Dr. Stein Aerts, and Dr. Piero Carninci for taking the time to evaluate my thesis critically and for the insightful discussions we had during the defense. I am incredibly thankful to Piero and Charles Plessy for motivating me to pursue the PhD in Bart's lab.

I would also like to thank the whole lab, which has been like a family to me in the past five years. Thank you, Julie, for all the help. Thank you, Roel and Michael, for helping me out when I was lost and for all the time in Satelite. Thank you, Petra, Masha, Wouter, Pernille, and Vincent, for your kindness and for helping me figure out limitations in my reckless optimism. I want to thank three adipo-princesses of the lab, Magda, Julie, Rachana, and Radiana. Thank you for all the help and kindness in the past years. Thank you Daniel for teaching me to enjoy borscht and vodka and all discussions about photography. Thank you, Gerrard, for teaching me what VCM is and for teaching me to value importance of sports in wellbeing. At the end, this thesis was just another race. I owe my special gratitude to the band of brothers – the techno part of the lab. Johannes (Mr. Bus and fellow sufferer), Wanze (come to the center), Jörn (Jörnie boi), and Riccardo (Riccito), thank you for challenging me to think outside the realm of classic molecular biology, mentorship, your patience, crazy experiments, starcraft and beer sessions, and all the time we shared in the past years. Time spent with you made this journey exciting and worth every moment spent in this laboratory.

I want to thank Aleksandar, Aleksandra, and Aleksandra for being my triple A batteries outside the lab and for introducing me to the beauties of Lausanne and Switzerland. I would also to thank many others I have spent my time in Lausanne, especially the Route de Vallaire crew and to all people who supported me from Belgrade (and all over the world), especially Molekolozi and Spakoglavica fan club members. On that note; I am incredibly thankful for my

two godfathers, Djordje (malogradjanski kum) and Lazar (romski kum). I shared my darkest and happiest moments in the past decade with those two mad lads. We ended up in different parts of the world, yet it felt like we were still together like during our studies.

I am eternally grateful to my family, Bratislave, Teodore, Nataša, for being supportive and believing in me sometimes more than I did in the past years. Miroslava and Danilo, thank you for your unconditional love during the chaotic times. Milovane and Slaviša, thank you for training my curiosity, being the voice of reason, and supporting this crazy path. This pack of wolves is responsible for who I am today and taught me how to move through life.

Finally, Melody, thank you for being my partner in crime, for your direct support, your kindness, love, patience, and for reminding me to keep my head up no matter what.

Thank you!

Abstract

Single-cell transcriptomics (scRNA-seq) started a technological revolution in biology by enabling through plethora of methods to assess a molecular state of the cell on systems level without strict necessity of the prior knowledge of the cell state. This wealth of data enabled us to answer fundamental problems of biology – how cells develop into plastic tissue, how they orchestrate developmental and homeostatic processes?

This thesis will cover developing novel methods and assays to study, map and perturb cell fates using droplet microfluidics based scRNA-seq. First part will cover our re-engineering Drop-seq setup, for broader adoption and more efficient sample handling. Developments presented will cover novel flexible workflow that significantly simplifies the sample handling while being efficient in single-cell material recovery.

Building on previous advances, next part will cover a novel dropleting platform for deterministic mRNA capture bead and cell co-encapsulation (DisCo) and its application to study single intestinal organoid development. Here, a key advancement, so far unsolved in the field is ability to deterministically control the cell capture process. Although at lower throughput than other droplet microfluidics-based methods, this method allows to process low cell input samples with high capture efficiencies. As a proof-of-concept study, we have studied early development of single intestinal organoids. DisCo helped us in uncovering extensive organoid heterogeneity and to identify the three distinct and aberrant subtypes of intestinal organoids such as enterocysts, spheroids and gobloids.

Finally, third part will cover development of novel gain of function barcoded transcription factor (TF) screening in conjugation with scRNA-seq (TF-seq). To understand role of TF in maintaining cell types, systematic manipulation of TFs combined by global gene expression profiling will aid in delineating their role in cell state progression and maintenance. We applied TF-seq on murine mesenchymal stem cell like C3H10T1/2 model and found multiple known factors that program cells to adipose (*Cebpa, Pparg*), osteogenic (*Runx2, Dlx3*), and myogenic lineages (*Myod, Myog*). Additionally, we discovered and confirmed two novel TFs (*Rhox12, Mycn*) involved in adipogenesis. By analyzing both TF activity and genome-wide changes caused, this approach has power to uncover both strong and weak drivers of lineage commitment among TFs.

Key words: Single-cell transcriptomics, droplet microfluidics, determinism, transcription factor, lentiviral vectors, mesenchymal stem cells, lineage commitment, differentiation

Résumé

La transcriptomique unicellulaire (scRNA-seq) a déclenché une révolution technologique en biologie en permettant, grâce à une pléthore de méthodes, d'évaluer l'état moléculaire de la cellule au niveau des systèmes sans qu'il soit strictement nécessaire de connaître au préalable l'état de la cellule. Cette richesse de données nous a permis de répondre à des problèmes fondamentaux de la biologie - comment les cellules se développent en tissu plastique, comment elles orchestrent les processus de développement et d'homéostasie ?
Cette thèse portera sur le développement de nouvelles méthodes et essais pour étudier, cartographier et perturber les destins cellulaires en utilisant des microfluidiques de gouttelettes basées sur le scRNA-seq. La première partie portera sur la réingénierie de notre dispositif Drop-seq, en vue d'une adoption plus large et d'une manipulation plus efficace des échantillons. Les développements présentés porteront sur un nouveau flux de travail flexible qui simplifie considérablement la manipulation des échantillons tout en étant efficace dans la récupération de matériel unicellulaire.

En s'appuyant sur les avancées précédentes, la prochaine partie portera sur une nouvelle plateforme de goutte-à-goutte pour la capture déterministe de l'ARNm par des billes et la coencapsulation cellulaire (DisCo) et son application à l'étude du développement d'organoïdes intestinaux uniques. Ici, une avancée clé, jusqu'ici non résolue dans le domaine, est la capacité à contrôler de manière déterministe le processus de capture cellulaire. Bien qu'à un débit inférieur à celui d'autres méthodes basées sur la microfluidique des gouttelettes, cette méthode permet de traiter des échantillons à faible apport cellulaire avec une efficacité de capture élevée. Comme étude de preuve de concept, nous avons étudié le développement précoce d'organoïdes intestinaux simples. DisCo nous a aidé à découvrir une grande hétérogénéité des organoïdes et à identifier les trois sous-types distincts et aberrants d'organoïdes intestinaux tels que les entérocystes, les sphéroïdes et les gobloïdes.

Enfin, la troisième partie portera sur le développement d'un nouveau criblage de facteurs de transcription (TF) à code barres de gain de fonction en conjugaison avec le scRNA-seq (TF-seq). Pour comprendre le rôle du FT dans le maintien des types de cellules, la manipulation systématique des FT combinés par le profilage global de l'expression des gènes aidera à délimiter leur rôle dans la progression et le maintien de l'état des cellules. Nous avons appliqué la TF-seq sur des cellules souches mésenchymateuses murines comme le modèle C3H10T1/2 et avons trouvé de multiples facteurs connus qui programment les cellules pour qu'elles deviennent adipeuses (*Cebpa, Pparg*), ostéogéniques (*Runx2, Dlx3*) et myogéniques (*Myod, Myog*). En outre, nous avons découvert et confirmé deux nouveaux FT (*Rhox12, Mycn*) impliqués dans l'adipogenèse. En analysant à la fois l'activité des TF et les changements causés à l'ensemble du génome, cette approche permet de découvrir les facteurs forts et faibles de l'engagement des TF dans les lignées.

Mots clés : Transcriptomique unicellulaire, microfluidique en gouttelettes, déterminisme, facteur de transcription, vecteurs lentiviraux, cellules souches mésenchymateuses, engagement dans la lignée, différenciation

Сажетак

Транксирптомика на нивоу појединачних ћелија (scRNA-seq) је револуционизовала молекуларну биологију кроз развој метода које су омогућиле системску анализу ћелија отклањајући потребу за предзнањем ћелијског статуса или идентитета. Подаци добијени применом ових метода нам омогућавају да одговоримо на фудаментална питања биологије: како се ћелије развијају у ткиво, како регулишу процесе развоја и хомеостазе?

Применом метода микрофлуидике и транскриптомике на нивоу појединачних ћелија, у овој докторској дисертацији ћу представити развој нових метода за изучавање, мапирање и манипулацију ћелија. Најпре ћу представити оптимизацију методе за узорковање ћелија базиране на микрофлудици (Drop-seq) са циљем побољшања њене ефикасности и ширег усвајања у научној заједници. Оптимизације методе покривају нови оптимизован протокол секвенцирања библиотика који поједностављa третирање узорака и ефикасно узорковање материјала за даљу анализу.

Применом принципа претходног поглавља, наредно поглавље покрива развој нове платформе за детерминистичку енкапсулацију куглица обележених баркодовима и ћелија (DisCo) како би изучавали развој појединачних интестиналних органоида. Развојем ове методе, омогућено је ефикасно узорковање и припрема библиотека за секвенцирање ретких и малих популација ћелија. Популације ових ћелија су до сада биле недоступне за анализу услед неефикасности до сада развијених метода. Како бисмо доказали значај ове технологије, проучавали смо развој појединачних интестиналних органоида. Развој ове методе нам је омогућио да опишемо хетерогеност у развоју органоида. Такође, овом анализом уочили смо 3 аберантна типа органоида које карактерише доминација терминално диференцираних ћелијских типова присутних и у дигестивном систему (ентероцисти, сфероиди, гоблоиди).

У трећем делу дисертације, представићу развој новог есеја за системску анализу функције баркодираних транскрипционих фактора на нивоу појединачних ћелија. У циљу бољег разумевања улоге транскрипционих фактора у дефинисању идентитета ћелије, систематска манипулација њихове функције у комбинацији са глобалном анализом експресије гена има потенцијал да открије њихову улогу у прогресији и одржавању стабилног ћелијског типа. Применом поменуте методе на модел мишијих мезенхимских матичних ћелија описали смо неколико до сада познатих транкрипционих фактора укључених у адипогенезу (*Cebpa, Pparg*), остеогенезу (*Runx2, Dlx3*) и миогенезу (*Myod, Myog*). Поред тога, описали смо два нова транкскрипциона фактора који регулишу адипогенезу (*Rhox12, Mycn*). Анализом утицаја транкскрипционих фактора на глобалну експресију гена омогућено је даље октривање нових , како јаких тако и слабих, транскрипционих фактора одговорних за програм диференцијације у специфичне типове ћелија.

Кључне речи: транкскриптомика, микрофлуидика, детерминизам, транскрипциони фактори, лентивируси, мезенхимске матичне ћелије, диференцијација

Table of contents

Acknowledgments 2

Abstract 4

Résumé 6

Сажетак 8

Table of Contents 10

Table of Figures 12

Chapter 1: Introduction 13

1.1 How does cell variation arise and what is its relevance? 13

1.2 The single-cell transcriptomics revolution 14

1.3 Advances and limitations in single-cell transcriptomics method development 16

1.4 Transcription factors and their role in developing and maintaining cellular identity 17

1.5 How can we study cellular heterogeneity using perturbation methods? 19

Chapter 2: Simplified Drop-seq workflow with minimized bead loss using a bead capture and processing microfluidic chip 22

2.1 Abstract 23

2.2 Introduction 23

2.3 Results 26

2.4 Discussion 35

2.5 Materials and Methods 36

2.5 Supplementary Figures 41

Chapter 3: Deterministic scRNA-seq of individual intestinal organoids reveals new subtypes and coexisting distinct stem cell pools 42

3.1 Abstract 43

3.2. Introduction 43

3.3 Results 46

3.4 Discussion 56

3.5 Acknowledgements 58

3.7 Materials and Methods 60

3.8 Supplementary Figures 75

3.9 Supplementary Table 79

Chapter 4: Single-cell RNA-seq guided transcription factor gain of function screen in mesenchymal stem cells 80

4.1 Abstract 81

4.2 Introduction 81

4.3 Results 83

4.4 Discussion 89

4.5 Materials and Methods 90
4.6 Supplementary Figures.. 97
Chapter 5: Discussion and Outlook.. 100
5.1 Method development in single-cell transcriptomics domain 100
5.2 Single-cell perturbation assays 102
5.3 Future perspectives 103
References.. 105
List of abbreviations.. 117

Table of figures

Figure 2.1 - Overview of the revised Drop-seq workflow 26
Figure 2.2 - The e-chip simplifies the experimental setup and renders it more flexible 27
Figure 2.3 - The novel bead capture and processing cp-chip increases the bead recovery efficiency 29
Figure 2.4 - qPCR species-mixing quantifies cross-contamination for different de-emulsification approaches 31
Figure 2.5 - STAMP generation on-chip 33
Supplementary Figure 2.1 - Impact of bead washing on background 41
Figure 3.1 - Overview and critical feature assessment of the deterministic co-encapsulation (DisCo) system 46
Figure 3.2 - Utilizing DisCo to map intestinal organoid cell heterogeneity along development 50
Figure 3.3 - Cell type distribution and marker gene expression across individual intestinal organoids during development: 54
Supplementary Figure 3.1 Schematic of the DisCo workflow and performance metrics...... 75
Supplementary Figure 3.2 - Quality assessment of the scRNA-seq data generated using DisCo from intestinal organoids 77
Supplementary Figure 3.3 - Individual intestinal organoid diversity 78
Figure 4.1 - Overview of theTF-seq principle... 83
Figure 4.2 - TF-ID enrichment overview 84
Figure 4.3 - Definition and characterization of functional cells in TF-seq framework............ 86
Figure 4.4 - Confirmation of TF activity and relationship to vector expression 88
Supplementary Figure 4.1 - Quality parameters of individual scRNA-seq experiments 97
Supplementary Figure 4.2 - Cell pre-filtering and functional cells with significant TF overexpression.. 98

Chapter 1: Introduction

1.1 How does cell variation arise and what is its relevance?

"A grand and almost untrodden field of inquiry will be opened, on the causes and laws of variation, on correlation of growth, on the effects of use and disuse, on the direct action of external conditions, and so forth. The study of domestic productions will rise immensely in value. A new variety raised by man will be a far more important and interesting subject for study than one more species added to the infinitude of already recorded species. Our classifications will come to be, as far as they can be so made, genealogies; and will then truly give what may be called the plan of creation" Charles Darwin, On the Origin of Species

Central dogma of molecular biology provides a simple yet powerful basis how genetic information is encoded and sequentially transferred in biological systems[1]. DNA is transcribed into messenger RNA (mRNA) which serves as an intermediary for a protein synthesis. However, due to its simplistic nature, the central dogma fails to unveil the complexity of the information flow, its divergence. This explanation also fails to deliver how complex biological entities emerged throughout the evolution and how such a rich phenotypical and functional diversity is achieved through organismal development. Additional question remains, how does the organizational diversity of biological systems arise from the same information source and how is it regulated[2].

The extent of complexity behind this flow of information helped us to understand how first unicellular lifeforms emerged and how their genome regulation transformation and cellular interactions allowed for the emergence of multicellular organisms and specific cell types[3]. Cells are the building blocks of multicellular organisms. A unique aspect point of metazoan evolution is tissue specialization and cellular function. This applies to both parenchymal cells that determine the tissue function on the one side and accessory cells that support and regulate their function on the other[4–6]. For decades, cell types were defined by their morphology and by their function [7,8]. With the advancement of microscopy methods, staining methods, and other techniques (i.e., flow cytometry), resolution and throughput of analysis increased, still leaving the possibility to probe a relatively low number of molecular and morphological features per single cell. An essential aspect of the cell type is its regulatory independence from the other cells. Without a unique regulatory program that enables cell-type-specific gene expression, it is hard to establish a cell-type identity. The selection of these features can lead to the narrow definition of cell identity reduced to several parameters. Numerous projects tried to address this question, such as FANTOM project[9], and ENCODE project[10]. Data generated by these projects allowed to analyze cells from a molecular perspective unveiling regulatory complexity behind cellular identity. With the rise of genome-wide analysis, scientists were able to pave a way towards broader measurements allowing for less biased cellular characteristics assessment. However, most of the mentioned methods still suffer from bulk data collection bias leading to the inability to resolve cell-to-cell differences.

1.2 The single-cell transcriptomics revolution

Single-cell RNA-seq (scRNA-seq) represents the array of methods developed in the last decade that allows mapping and dissecting cellular diversity[11,12]. The revolution in single-cell genomics lays it is foundation in the everlasting interest of scientists to systematically understand gene expression and underlying mechanisms behind cellular composition and behavior. From a molecular biology method development standpoint, the major breakthroughs happened with first development discovery of reverse transcriptase [13] and reverse transcription combined with PCR [14]allowing to measure activity of genes from cells. With development of PCR-based methods and second generation of sequencing scientists were able to increase and parallelize the gene coverage leading to development of various transcriptomic methods that enabled genome-wide assaying of the cellular and tissue activitiy[15,16].

Building on these principles, Tang and colleagues further developed and applied transcriptomics approaches to analyze the gene expression of single cells[17]. Followed by this, accuracy and sensitivity of the methods was on the rise in the last decade[18]. There are several notable waves of development that allowed for increase of experimental scale. First, integration of transcript barcoding [19] with precision of Fluorescence Activated Cell Sorting (FACS) technology allowed for the automated single-cell sorting directly into lysis buffers and integration standard molecular biology workflows (i.e. PCR, nucleic acid purification etc.) [20]. With further automation of sorting methods and biochemical reaction miniaturization, various liquid handling platforms were integrated with already established scRNA-seq protocols[21] .

Further increased sensitivity was achieved with development of microfluidics methods that enabled further miniaturization of RNA capture and reverse transcription [22–24]. Droplet-based microfluidics is a set of techniques commonly used to perform experiments in the nanoliter scales. The main advantage of such approach is rapid compartmentalization of water-in-oil droplets that allows for reactions to happen fast, on a smaller scale with high resolution[25–27]. Hence, droplets give ability isolate particles of interest and perform the reactions in temporal and spatial resolution. Ultra-high throughput nature of these methods (reference for throughput) made them attractive to perform various screening assays in life sciences such as fluorescence sorting[28], antibody screening[29] and cell culturing[26,30]. After 2015, together with development of microsphere oligo synthesis coating, droplet microfluidics revolutionized single-cell transcriptomics field allowing rapid cell compartmentalization and barcoding in the individual droplets[31–35].

We were highly encouraged in our decision to implement a novel single-cell method, Drop-seq [32] that exploits high throughput droplet microfluidics and unique bead barcoding to obtain single-cell expression profiles [32,36,37]. Drop-seq was preferable platform since it suffers from much fewer cell doublets and is much more cost-effective to reach desired power for gene expression analysis because it allows the processing of thousands of single cells in one run with single-cell purity between 90 and 98% [32,38]. For example, compared to other trap-based microfluidics platforms like Fluidigm C1, droplet microfluidics doesn’t suffer from cell selection bias, important for heterogeneous and tissue samples, thus doesn’t suffer high doublet rates[32,39,40]. Compared to other Fluorescence-Activated Cell Sorting (FACS) based methods, Drop-seq does not suffer from limitation to 96 or 384 well format sample processing and allows more efficient sample size scaling[41,42]. Once collected, thanks to unique bead barcoding, single cells in the Drop-seq experiment are processed in bulk, making each experiment less labor-intensive and free from necessity to utilize robotic liquid handling systems. Commercial release of the droplet-based technology by company 10X Genomics significantly improved the throughput and efficiency of single-cell capture[43] in droplet microfluidics and paved a way through method standardization and ease-of-use towards broad adoption of this approach.

A Drop-seq experiment begins with the encapsulation of cells and beads in nanoliter water-in-oil droplets. The system relies on unique barcoding of the beads co-encapsulated with cells. The beads are coated with uniquely barcoded primers, each of which contains a cellular barcode (specific for every bead), a unique molecular identifier (UMI, specific for each primer) and a polydT sequence used for polyadenylated mRNA capture. Before injection, beads are suspended in a lysis buffer that, upon cell encapsulation, induces immediate mRNA release. The system relies on stochastic co-encapsulation with 2-10% single-cell capture efficiency [32,42,44]. When cell/bead co-encapsulation occurs, the beads capture released mRNA. The newly formed bead/RNA complexes are called STAMPs (Single-cell Transcriptome Attached to Micro Particles). After formation, droplets containing STAMPs are pooled down together in a single tube. At this point, STAMPs are centrifuged and suspended in a reverse transcription (RT) mix used to convert captured RNA into stable complementary DNA (cDNA). After RT, STAMPs are treated with exonuclease I to remove unbound oligonucleotides on beads. Following this step, cDNA amplified directly from the beads, then the sample pool is prepared for paired-end sequencing[32] .

1.3 Advances and limitations in single-cell transcriptomics method development

At this moment, single-cell RNA-sequencing (scRNA-seq) has become an indispensable tool in biomedical research by enabling the exploration of cell-to-cell differences in tissues at unprecedented resolution[45–47]. With single-cell approaches as an asset, one can fully characterize the functional diversity of the tissue, its maintenance, and remodeling caused by environmental perturbations. Data generated from the single-cell experiments provides an opportunity to understand transcriptional variation and cell-to-cell heterogeneity arising from the mutations, cell cycle, cell-to-cell communication, and various subpopulations present. Finally, cell type identity definition has evolved and can be defined through understanding four aspects – phenotype, function, lineage, and state[48,49]:

Each one of these parameters can help us to pinpoint cellular identity in a less biased manner. Current experimental and computational advances in scRNA-seq have already revealed and confirmed numerous insights into various biological systems. Following the wave of standardization among the methods[50,51], both on bench and computational side, and with increase interest increased interest among scientist to perform unbiased studies of thousands of single cells within a sample [41,52], novel consortia emerged with a common goal – to fully catalog all existing cells in relevant model organisms[11,53–55].

Many single-cell transcriptomics methods required already wide-spread skills and standard equipment that led to their rapid adoption. On the one side, methods that rely on index sorting were developed around existing sorting equipment and built on extensive miniaturization (nanowells) and automation (robotic liquid handling systems). On the other side, as adoption of microfluidics techniques increased by various labs, novel microfluidic approaches laid ground for broad adoption and innovation in biology. However, certain limits in the current method development are still present: loss of information due to inefficient RNA to cDNA conversion (commonly called as dropouts) still significant, sensitivity in detection varies depending on the chemistry of choice, majority of methods suffer from bias to polyA capturing leaving some regulatory parts of transcriptome in dark [41,52]. For latter, several advances were recently made, however, such as methods one still has to choose between "unbiased" but in fact polyA based transcriptome or methods focus on other RNA species within of the transcriptome[56] .

While available scRNA-seq technologies allow medium- to high-throughput routine experiments, efficient sample processing is still a challenge for all approaches. Multiplexing techniques were developed to permit parallelization of sample analysis for efficiency and cost-effective reasons. Still, current multiplexing techniques (i.e., CITE-seq, MULTI-seq)[57,58] require several magnitudes higher starting material to compensate for all the sample losses due to washing steps and do not address losses before multiplexing (i.e., tissue digestion before the labeling). As another alternative, plate-based assays have made enormous advances in ease-of-use and are continuously evolving, however, they suffer from low cell capture efficiency. FACS [23]as being the one that is convenient and well established for molecular biology labs [59]minimally 10,000 input cells . Furthermore, sophisticated automated liquid handling, also not available for most labs, would be required.

1.4 Transcription factors and their role in developing and maintaining cellular identity

One of the main gatekeepers and regulators of genome activity are transcription factors (TFs). Transcription factors serve as chain-links of the genome to other functional units of the cells through DNA binding and recruiting of RNA polymerase complex and other cofactors. Majority of the genome's functional DNA is likely regulatory and TFs play a central role in its recognition and utilization. TF activity controls processes that specify cell types, developmental patterning, and specific pathways that govern tissue plasticity and homeostasis. TFs differ in their impact on transcription upon DNA binding. The binding of TFs to DNA determines the strength and time-window of gene regulation. Some TFs can directly recruit RNA polymerase while others recruit additional factors that promote specific transcription phases TFs can also be classified as either “activators” or “repressors”. Activators do… Repressors do… However, this classification can be easily blurred[60] as some TFs recruit multiple cofactors that have opposing effects depending on the local sequence context, availability of cofactors and dosage[61,62]. Name an example Is there any other way of classifying them?

Approximately 1500 TFs are encoded in the mammalian genome and are associated with various diseases and phenotypes[63]. How TFs find their targets and control gene expression patterns are essential for understanding both homeostatic physiological functions and pathologies. One of the first examples *Myod* shown to a critical TF for commitment and differentiation to muscle cells. Previous studies show that both *Myog* exert their function by regulating genes involved in the myogenic program through promoter-specific regulation and positive-feedback[64].

Across all TF families, many are tissue-specific, thus maintain specific physiological functions. Tissue specificity is enabled by temporal-spatial gene expression patterns driven by TFs. The nature of reprogramming reflects the functions of the TFs role in determining the cell fate. High-resolution data can explain associations between TFs, downstream genes regulated and ultimately cell identity. Decades of research have made significant progress in understanding the roles of TFs in various biological systems, yet we have recently started to understand the genome-wide causative relationships between TFs and cell states[65,66]. To further aid the identification of cell identity, it is important to analyze further the structure and function of TF regulatory networks[67]. First studies using chromatin immunoprecipitation and sequencing (ChIP-seq), yeast one hybrid[68] enabled to identify TFs and their interaction partners. Methods like[69], mechanically induced trapping of molecular interactions (MITOMI)[70], and systematic evolution of ligands by exponential enrichment (SELEX) enabled us to further expand our knowledge on binding affinity, frequency, and binding mode of TFs. Additionally, Live-cell imaging methods allowed for closer tracking of TFs dynamics throughout the cell cycle and cell-fate decisions allowing for further dissection of TF activity in active cells[71]. Furthermore, gene perturbation experiments, both gain-of-function and loss-of-function, are increasingly complemented with genome-wide analysis to study their importance in defining cellular phenotype, function and lineage progression of cells.

1.5 How can we study cellular heterogeneity using perturbation methods?

Understanding underlying molecular mechanisms during cell differentiation in a multitude states and types has been a long-standing interest in stem cell and developmental physiology. This knowledge has a propensity to unveil how to manipulate cell fates *in vivo* and to understand the origins of different pathologies during organismal development and aging. Stem cells are capable of self-renewal and to differentiate to specific cell types. Their ability to maintain and use these two parameters depends on several parameters: the niche, the epigenetic control, and the transcriptional control[49,72]. However, the emergence of the novel *in vitro* techniques allows for manipulating epigenetic landscape and transcriptomes, enabling efficient studies and trans-differentiation to novel cell states[73].

Revolution in scRNA-seq also allowed for the development of screens with single-cell readout, which delivers a way for dissecting complex outputs not easily reduced to single markers. This allows for initial dissection of complex signaling pathways and other biological mechanisms under perturbation conditions on an unprecedented resolution and understanding and continuity of cell states along the developmental trajectory progressing towards a terminally differentiated state[74]. Classical studies that involve the study role of genes that drive the reprogramming of cells have heavily relied on prior knowledge of their role. Cellular reprogramming via gene manipulation has widely impacted biological research from reprogramming adult parenchymal cells to the induction of pluripotent stem cells[75].

There are three most notable approaches to screening of perturbations in stem cells. First, methods like CellTagging and LARRY, enable combinatorial cell labeling in order to reconstruct lineage trees[48,74]. Second, CRISPR/Cas screens coupling with scRNA-seq have performed multi-locus gene perturbation, focusing on the TFs regulating immune response [76–78]. This approach is known as Perturb-seq and gives us the possibility to use the CRISPR/Cas system to target specific genes to assess the diversity of transcriptional changes and gene interactions across thousands of knockout cells. Similarly, Nguyen *et al.* have generated large transcriptomic dataset compromising CRISPRi human induced pluripotent stem cell (hiPSC) line [79]. This allowed them to identify cells' subpopulations based on their regulatory networks controlling the identity and pluripotency differentiation potential. CRISPR offers advantages such as scaling up, targeting, and the ability to mimic endogenous activation. However, overexpression of TFs offers a stable expression of the gene of interest, enables manipulation of specific isoforms and mutant forms, which cannot be easily probed by endogenous activation[80,81].

One of the Laboratory of Systems Biology and Genetics' longstanding interests is the adipose tissue and its homeostasis[82,83]. It is now well appreciated in the scientific community that adipose tissue is not an expendable part of our body but constitutes a highly dynamic, endocrine-type organ playing an indispensable role in several biological processes, including development, immunity, and energy homeostasis [84,85] . The adipose tissue can be classified into three subtypes based on their molecular and morphological characteristics. The best known and the most abundant is the white adipose tissue (WAT). This tissue is characterized by large cells that store energy in the form of unilocular fat droplets. Moreover, it is now well established that both in mice and humans, WAT depots are moleculary, cellularly and physiologically distinct depending on their location in the body which may reflect the fact that these adipose depots stem from different fat precursors[86,87]. In contrast, the lesser-known brown adipose tissue (BAT), which has recently been shown to be present in both human infants and adults[88], contains cells with multilocular fat droplets and serves to expend energy as heat. Finally, "beige" adipocytes [89] exhibit a brown fat cell-like phenotype. Beige adipocytes are thought to be ontogenetically and molecularly distinct from brown adipocytes given that they arise in WAT upon extensive exposure to cold, exercise or adrenergic stimuli[90]. They have distinct gene expression profiles, and they are thought to have a different developmental origin than genuine brown fat cells[89,91].

Adipocytes are derived from mesenchymal stem cells (MSCs) in a two-step process[92], including the determinant differentiation step from MSC to pre-adipocyte and terminal differentiation step from pre-ad to mature adipocyte[93]**.** To study terminal differentiation of adipocytes, Gubelmann *et al.* transduced individual 734 mouse TFs packaged as lentiviruses in 3T3-L1 cells in the previously established overexpression screen. Their screen showed 26 TFs that significantly induce the white adipogenesis, out of which 23 were already known. However, using a plethora of other less well-characterized TFs can reveal many novel regulators of these differentiation processes. Although the TFs that control the terminal differentiation are well studied, much less is known for the determinant differentiation. Lack of solid knowledge in this subdomain is due to the fact that pre-adipocyte transient state harbors neither specific markers nor show a significant phenotype. Our lab possesses a library containing most mouse TFs (~734) in lentiviral expression-ready vectors, which constitute an indispensable resource for further TF gain-of-function studies[94].

Chapter 2: Simplified Drop-seq workflow with minimized bead loss using a bead capture and processing microfluidic chip

Marjan Biočanin[124] Johannes Bues [124] Riccardo Dainese,[12] Esther Amstad [3] and Bart Deplancke 12

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, Lausanne, Switzerland

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

4 Contributed equally to this work.

Shared First Author Contributions: MB and JB conceptualized the study, JB developed the cp-chip, MB developed the e-chip, JB developed the qPCR assay, MB and JB measured bead recovery efficiency, MB performed Drop-seq experiments, MB and JB analyzed the Drop-seq data, MB and JB wrote the manuscript.

Chapter information: This chapter is published in Lab on a Chip, DOI: 10.1039/c9lc00014c

2.1 Abstract

Single-cell RNA-sequencing (scRNA-seq) has revolutionized biomedical research by enabling the in-depth analysis of cell-to-cell heterogeneity of tissues with unprecedented resolution. One of the catalyzing technologies is single cell droplet microfluidics, which has massively increased the overall cell throughput, routinely allowing the analysis of thousands of cells per experiment at a relatively low cost. Among several existing droplet-based approaches, the Drop-seq platform has emerged as one of the most widely used systems. Yet, this has surprisingly not incentivized major refinements of the method, thus restricting any lab implementation to the original Drop-seq setup, which is known to suffer from up to 80% bead loss during the process. In this study, we present a systematic re-engineering and optimization of Drop-seq: first, we re-designed the original dropleting device to be compatible with both air-pressure systems and syringe pumps, thus increasing the overall flexibility of the platform. Second, we devised an accompanying chip for post-encapsulation bead processing, which simplifies and massively increases Drop-seq's cell processing efficiency. Taken together, the presented optimization efforts result in a more flexible and efficient Drop- seq version.

2.2 Introduction

The generation of high-dimensional gene expression data through single-cell RNA-sequencing (scRNA-seq) is becoming indispensable to explore the diversity of heterogeneous cell populations with the necessary accuracy and resolution[11]. For several of the catalyzing methodologies, microfluidics has quickly become a key investigative technique for cell capture, isolation and processing. The first, widely successful, micro- fluidic implementation of scRNA-seq was introduced and commercialized by Fluidigm, with its C1 system, which employed hydrodynamic traps for single cell capture and microchambers for downstream library preparation. While this approach allowed for convenient sample processing, costs per cell remained high, the throughput limited and doublets – i.e. the unwanted capture of two cells instead of one – was unexpectedly frequent[32].

With the emergence and maturation of droplet microfluidics in recent years, researchers have addressed the need for high throughput, low-cost assays by reducing reaction volumes[36]. Droplet microfluidics now allows for experiments on massive scales both within and outside of the “-omics” do- main[26,95]. Focusing on single cell transcriptomics, the possibility to perform rapid compartmentalization has significantly increased the throughput and sensitivity of analyses at the single cell level. For instance, recent publications contained 100 605 “Tabula muris” cells[96], 492 949 mouse nervous system cells[97], 157 000 ageing fly brain cells[55], and

10× genomics has publicly available datasets with 1.3 M cells in total[43,98]. While commercial systems such as the 10× genomics chromium exist, offering an indisputable ease of use and experimental robustness, customized lab setups are still a magnitude less expensive in terms of per-cell library preparation costs once established and routinely utilized.[42] Furthermore, custom- built setups offer the highest degree of flexibility, allowing for freely scalable experimental size and completely open source processing chemistry. Hence, while commercial systems are most arguably dominating the conventional cell pro- filing market, it is likely that, especially for power-users and research and development, custom-built setups will continue to play an important role[51,99–101]

The two most noteworthy custom-built dropleting systems for scRNA-seq are Drop-seq[32] and inDrop.[37] Conceptually, both systems are similar: a cell suspension is co-flown with a bead suspension, and encapsulated in nanoliter droplets. The beads harbor barcoded oligonucleotides for capturing polyadenylated RNA (polyA) and for adding a common molecular cell identifier to each mRNA transcript. In the bead suspension, detergent is added to lyse cells upon co-encapsulation and thus to liberate cellular polyA RNAs for subsequent capture. Beads for both systems, hydrogels for inDrops and microspheres for Drop-seq, are commercially available. Alternatively, inDrop beads can be produced in-house by split-and-pool synthesis[102], while the Drop-seq microspheres are more complex to manufacture since the process involves solid-phase oligonucleotide synthesis. Noteworthy, split-and-pool barcode synthesis has recently been demonstrated on solid-state microspheres[103], potentially making this an alternative route for Drop-seq bead generation. Despite having lower detection sensitivity based on artificial RNA molecules[52], and significantly lower cell capture efficiencies, Drop-seq still has found wide adoption in the scRNA-seq community[74,104–106]. This is because the utilization of solid-state (toyopearl) microspheres in Drop-seq, as compared to hydrogels in inDrops, makes Drop-seq relatively straightforward to implement. An additional technical advantage of using microspheres is that they can be directly coated by solid-phase synthesis, resulting in a high cell barcode diversity. Furthermore, the conceptually simple post-encapsulation reverse-transcription (RT) process makes the protocol comparably easy to establish: the RNA-harboring spheres can be retrieved from the emulsion, mRNA reverse-transcribed in a reaction vessel, and unhybridized bead-bound oligos removed by exonuclease I (Exo I) treatment.

However, despite its wide adoption, few improvements, especially for practical use of the system, have so far been described[106–108].Thus, if a lab aims to establish Drop-seq today, there is no alternative besides replicating the previously published approach, regardless of the equipment available in the lab or experimental needs. This implicitly forces for example air pressure-based microfluidics lab to purchase syringe pumps and a specialized magnetic stirring system. To this day, only Stephenson and colleagues[100] have built a low-cost, 3D printed pressure-based Drop-seq setup with instructions on how to specifically adapt the chip and operating equipment. Moreover, some experimental procedures, such as the post-encapsulation processing of the mRNA-bearing microspheres to form STAMPs (for Single-cell Transcriptomes Attached To MicroParticles), are still rudimentary: microspheres are handled in large vessels requiring a plethora of manual processing steps. While this not only introduces many error prone steps, up to 80% of all processed beads are lost during this process.[32] To our knowledge, no improvements have been made to the original bead processing protocol since its release.

To provide the community with a universally implementable system, we re-engineered the Drop-seq system and addressed one of its major limitations: bead loss. First, we show that our newly designed encapsulation chip runs robustly both on air- pressure and syringe pumps, thus providing greater experimental flexibility. Second, we show that the use of an air pressure source compared to syringe pumps allows for the integration of simple sample supply reservoirs, which alleviate the necessity of acquiring expensive micro stirrer setups. Third, we devised a novel capture and processing chip that allows for highly efficient retrieval of the mRNA capture beads. We demonstrate that this approach increases recovery efficiencies of micro- spheres, either from pre-broken emulsions, or directly from droplets. Using the same device, we show that the complete cDNA generation protocol can be executed on-chip, further streamlining the protocol. Finally, we confirmed that key scRNA-seq performance metrics are comparable to the original implementation of the protocol. Overall, our solution represents in our opinion a more versatile and robust approach compared to the original Drop-seq implementation.

2.3 Results

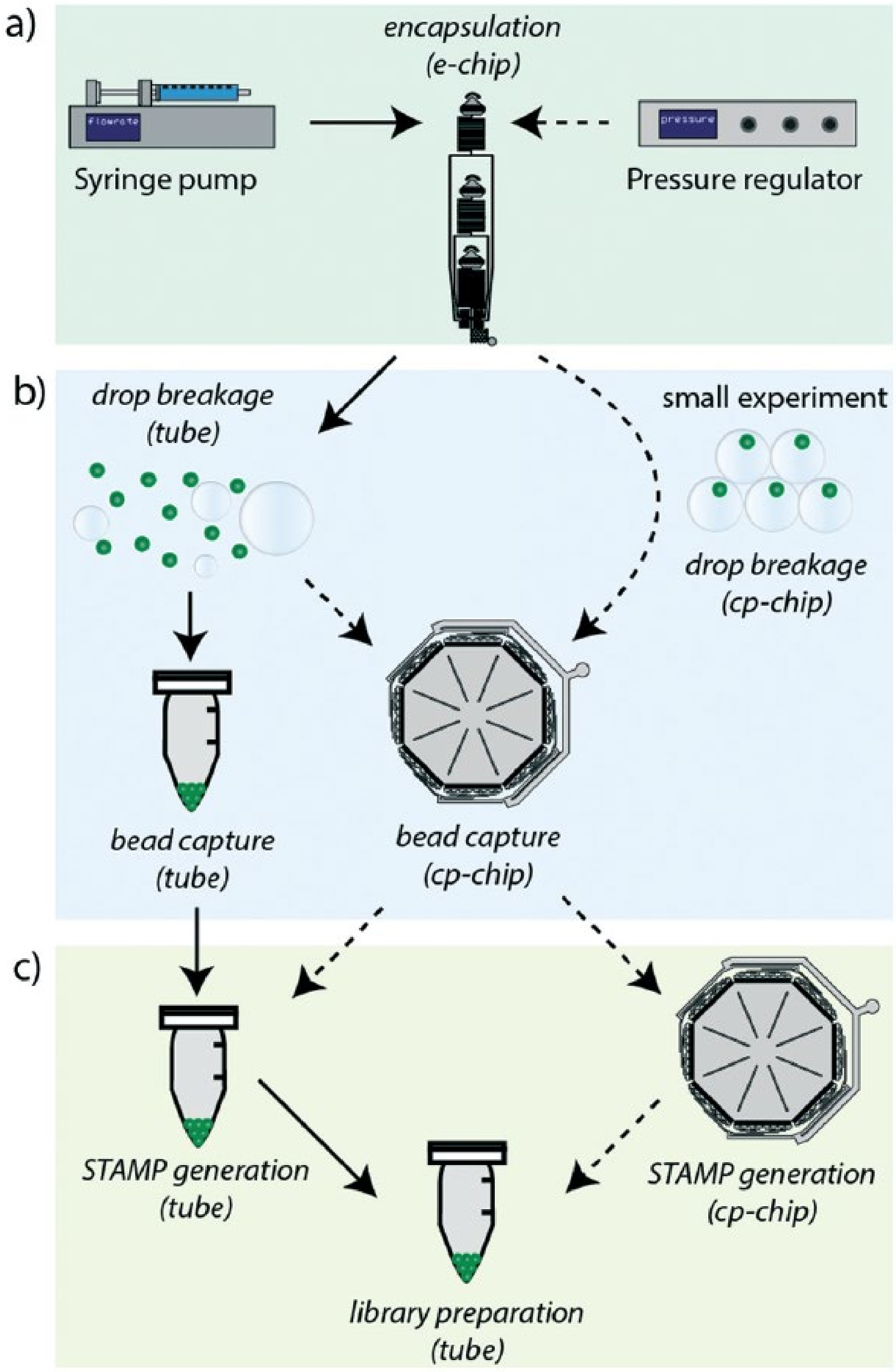

a)
encapsulation
(e-chip)
Syringe pump
flowrate
pressure
Pressure regulator
b)
drop breakage
(tube)
small experiment
drop breakage
(cp-chip)
bead capture
(tube)
bead capture
(cp-chip)
c)
STAMP generation
(tube)
STAMP generation
(cp-chip)
library preparation
(tube)

Figure 2.1 - Overview of the revised Drop-seq workflow: the revised work-flow (dashed lines) is a flexible protocol that simplifies handling and in-creases the efficiency of the original protocol (solid lines). (A) By redesigning the original chip into the e-chip, it is now possible to use either syringe pumps or an air-pressure regulator for the cell-bead encapsulation process. **(B)** Utilizing a new bead capture and processing chip, the cp-chip, beads can now be either captured post droplet breakage, or directly captured from droplets for smaller bead quantities. **(C)** Following the bead capture, it is possible to either retrieve the beads into a tube for the STAMP generation process (reverse transcription and exonuclease treatment), or to perform STAMP generation on-chip to further improve the overall bead recovery efficiency.

The revised Drop-seq workflow comprises two microfluidic devices, an encapsulation device (e-chip) and a bead capture and processing chip (cp-chip). The e-chip facilitates the cell bead encapsulation process and is compatible with pressure and syringe pumps (Figure 2.1A), whereas the cp-chip is employed for efficient bead extraction, either directly from droplets for small samples, or from broken emulsions for larger samples (Figure 2.1B). While it is then possible to proceed to cDNA generation in tubes after bead capture on the cp-chip by eluting the microspheres, it is also possible to generate STAMPs on chip (Figure 2.1C). The latter has the advantage of both minimizing bead loss and handling errors as well as offering the potential for automating this otherwise error-prone and labor- intensive procedure.

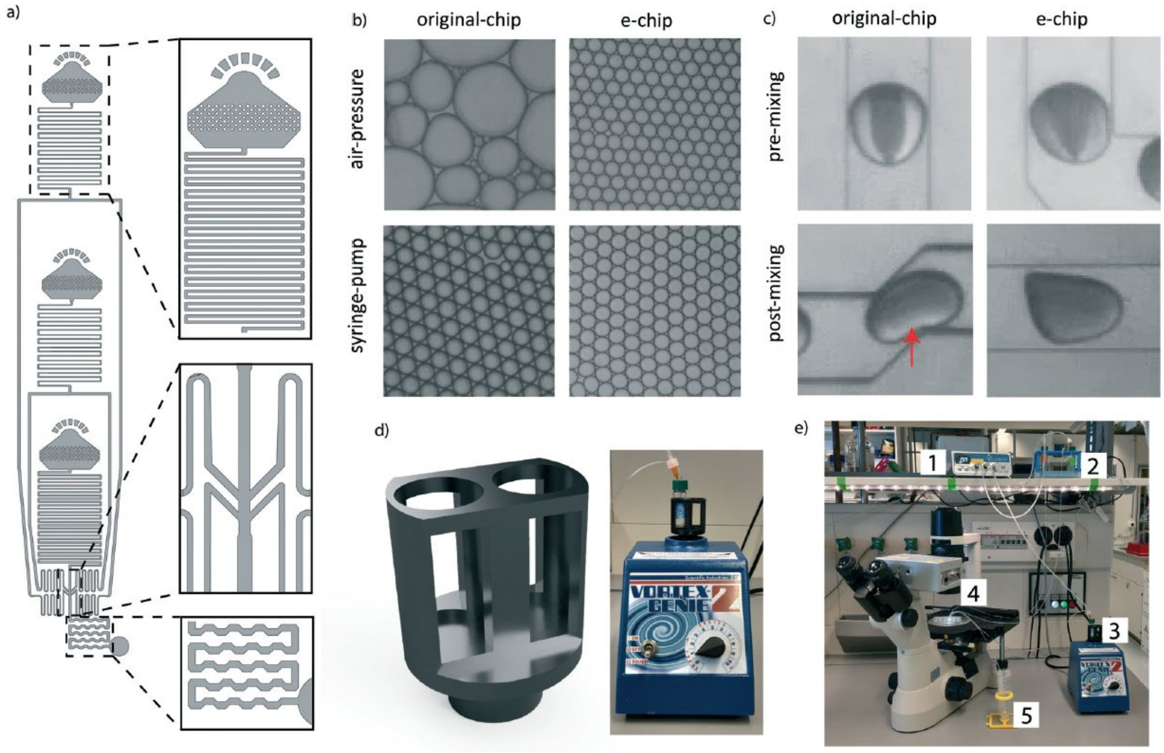


Figure 2.2 - The e-chip simplifies the experimental setup and renders it more flexible: (A) the e-chip includes flow resistors, an optimized flow focusing point, and extended droplet mixing channels (top to bottom). **(B)** Emulsion formation by the original device and e-chip using syringe pumps and an air-pressure source. **(C)** Droplets filled with transparent liquid and food-coloring after droplet formation (pre-mixing), and after exiting the droplet mixing channel (post-mixing). **(D)** Two-slot sample holder for bead suspension, attachable to a standard lab vortex machine. **(E)** Overall setup utilizing air-pressure as driver. An air pressure controller (1) is connected to the oil and cell reservoirs (2), and to the bead vessel (3). All vessels are connected to the dropleting device (4) and the emulsion is captured in a tube (5).

To make the workflow compatible with different pressure sources, a new encapsulation chip (e-chip) was designed to facilitate monodisperse emulsion formation with both air-pressure and syringe drivers (Figure 2.2A). Syringe pumps are fundamentally different from air-pressure systems as they are controlling volumetric flow rates. In contrast, air pressure systems only control the pressure of a liquid containing vessel. As all of these systems feature minimal operating pressures, the flow-resistance of the chip had to be adjusted to allow for appropriate flow rates. Thus, we first re-designed the original chip by adding microfluidic resistors to each channel, aiming to stabilize the flow rate in the operating range of the air pressure controller. These resistors were designed by considering the following criteria: a) The channel cross- section had to be sufficiently large to allow the easy passage of cells and beads and thus to avoid clogging. b) The resulting flow rate should not exceed the one used in the original Drop-seq (i.e. below 60 µL min−1). These criteria translated into 4.5 cm-long resistors with a squared cross-section with lengths of 60 µm. As the smaller channel width compared to the original device made the chip more sensitive to clogging, the second addition was large input filters to retain abnormally big microspheres and dust from entering the chip. Third, the dropleting T-junction was adapted into an angled conformation. We compared the performance of the newly developed e-chip to the original Drop-seq device by generating emulsions with both syringe pumps and pressure controllers (Figure 2.2B). While the original device clearly failed to produce monodisperse emulsions from a pressure controller, the e-chip was able to produce stable emulsions with both drivers. Importantly and independent of the driver system, we also observed that the length of the chaotic advection mixers on the original chip was insufficient to completely mix droplet contents (Figure 2.2C). Since passive mixing in droplets by diffusion is slow and uncontrollable, we decided to increase the length of the mixers on the e-chip four-fold. Indeed, a qualitative inspection confirmed that the larger mixing channels enhanced the internal mixing of the microdroplets (Figure 2.2C).

A positive side effect of using air pressure systems is the fact that conventional sample vessels can be utilized and interfaced with the chip, as opposed to syringes used in syringe pumps. This is especially significant for solutions containing microbeads, as they have to be agitated for the particles to remain in suspension. In the original Drop-seq implementation, this was achieved by an expensive magnetic stirrer setup (~$2000), connected in proximity of the syringes. To overcome the need for a dedicated magnetic stirrer, we developed a vessel agitation system that is compatible with conventional laboratory equipment. Using additive manufacturing, we produced a Vortex adapter that can hold and agitate the cell and bead suspension vessels (Figure 2.2D). We extensively used the Vortex-based agitation system for cells and beads and did not observe any adverse effect on bead integrity (data not shown). Besides simplifying the overall system (Figure 2.2E), this has the additional advantage of reducing the initial set-up costs.

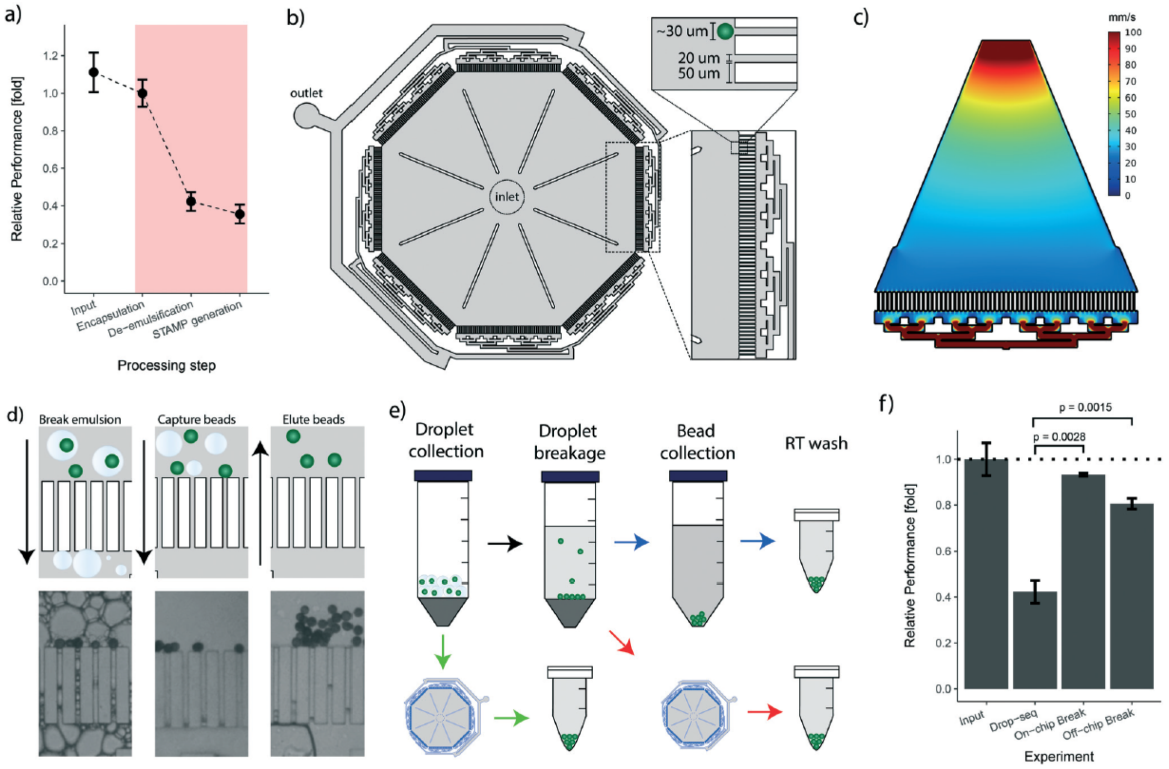


Figure 2.3 -The novel bead capture and processing cp-chip increases the bead recovery efficiency: (A) efficiency of the various bead processing steps for the original drop-seq protocol. Input amount is displayed relative to the encapsulation process. The efficiency of the de-emulsification step is normalized to the encapsulation step. The bead recovery efficiency for the STAMP generation process is shown relative to the de-emulsification process and normalized for 'sampling loss'. Error bars represent the standard deviation calculated on the normalized values for each processing step. **(B)** Design of the novel bead capture and processing cp-chip. **(C)** Hydrodynamic simulation of one chip section of the chip shown in (B). d) The device from (B) is capable of capturing beads directly from droplets (left) or from a broken emulsion (middle). Captured beads are eluted by reversing flow direction (right). **(E)** Different strategies for bead recovery from droplets: original protocol (blue arrows), collection on-chip from broken emulsion (red arrows), and bead capture from droplets (green arrows). **(F)** Bead recovery efficiencies for the different procedures shown in (E). Values are normalized to encapsulation losses; the error is shown as standard deviation calculated on the normalized values. P-values were deter-mined using t-test.

Following the cell encapsulation process, the original Drop- seq protocol involves a multistep procedure to recover the microspheres from the emulsion for subsequent STAMP generation. As these processing steps are conducted in vessels and equipment tailored to large volumes, they are naturally prone to bead loss. In the original protocol, this loss was quantified as ranging between 60–80%[32]. To identify critical steps during bead processing, we examined the bead recovery efficiency of the workflow by quantifying bead loss for each of the three main protocol steps, i.e.: encapsulation, de-emulsification, and STAMP generation (Figure 2.3A, details in the Materials and Methods section). The average number of input beads for efficiency measurements were 2249 ± 250 beads. As expected, we found that every manual handling step introduces bead loss (Figure 2.3A). Nevertheless, we observed that the steps executed on a microfluidic chip (encapsulation) and in small vessels (STAMP generation) show a much better performance (~10–20% loss) than the de-emulsification step performed in large vessels (~58% loss). Overall, we observed cumulative post-encapsulation bead losses of up to ~64%, which is in accordance with the reported efficiency of the most recent protocol.

To address the post-enapsulation inefficiencies that cause substantial bead loss, we designed a bead capture and processing chip (cp-chip) that simplifies the de-emulsification process (Figure 2.3B). Specifically, the cp-chip is comprised of pillar arrays that allow liquids to pass through while blocking particles above a certain size. Gaps between the pillar were 20 μm, substantially below the average size of the coated oligonucleotide beads (30 μm). The pillars were 50 μm wide, were arranged in eight linear sections of 4.1 mm each, and were combined in an octagonal shape to maximize the capture capacity of the device (Figure 2.3B). In principle, the cp-chip can hold more than 1000 beads in a single bead layer, and it can easily purify large Drop-seq samples of more than 10 000 beads in multi-sediment layers. To seed the array evenly across all sections, the channels towards the waste channel were arranged so that flow spreads across all pillars, con- firmed with a flow simulation (Figure 2.3C). Finally, large support structures were included to prevent collapse of the chip ceiling and to provide a guide for precise hole punching.

Initially, we tested the cp-chip qualitatively to demonstrate that the device was indeed capable of extracting beads from complex solutions. Intuitively, direct on-chip bead capture from emulsions represents the most straightforward approach as it involves no prior processing steps. However, since droplets that contain contaminating unbound mRNA are in direct proximity of beads coated with mRNA-capturing probes, this approach can potentially lead to increased cross-contamination of sequencing libraries. It was therefore important to validate that the cp-chip can capture beads that stem either directly from droplets or from the broken emulsion (Figure 2.3D). Next, we further examined both workflows (i.e. beads from droplets vs. broken emulsions) (Figure 2.3E) to determine their performance in terms of bead recovery efficiency compared to the original Drop-seq approach. To this end, beads were encapsulated in emulsions and recovered using each of the three different procedures. We found that de-emulsification using the cp-chip increased the recovery efficiency approximately two-fold compared to the original approach, a significant improvement (approximately 81% for bead capture from broken emulsions, and 93% for bead capture from droplets; Figure 2.3F).

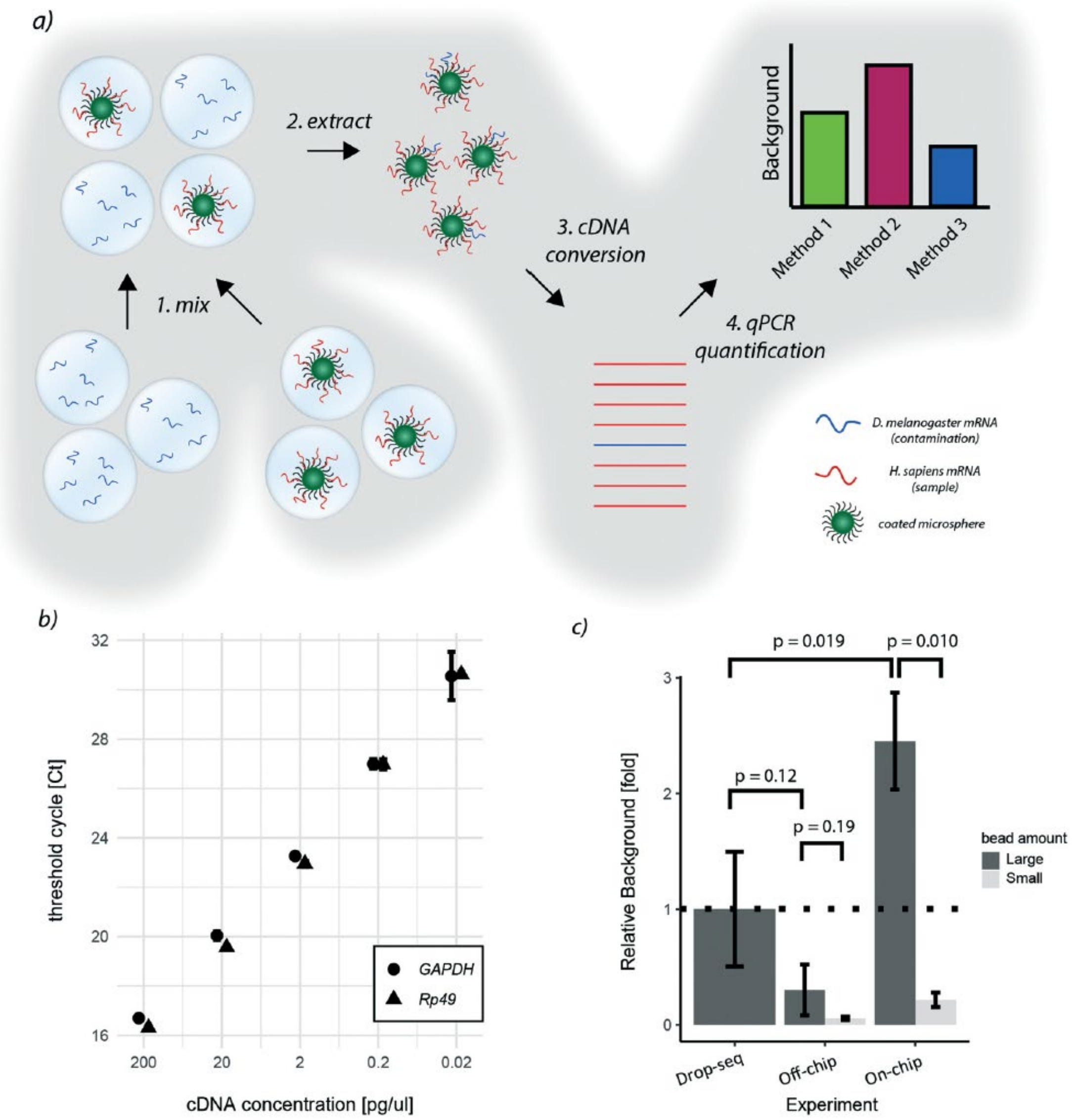


Figure 2.4 - qPCR species-mixing quantifies cross-contamination for different de-emulsification approaches: (A) qPCR approach to quantify species cross-contamination for de-emulsification strategies. Two emulsions were generated: one emulsion with droplets containing free floating *Drosophila* RNA, and a second emulsion containing microspheres hybridized with human RNA. Both emulsions were mixed, and beads extracted from the emulsion. Next, the captured mRNA was converted into cDNA and qPCR was performed with species-specific primers. Amplification cycles of each reaction allows for the quantification of cross-contamination. **(B)** Threshold cycles for the qPCR primers utilized for human cDNA (GAPDH) and *Drosophila* cDNA (Rp49). **(C)** Cross-contamination quantification for different de-emulsification strategies. Background values correspond to the ratio between Drosophila and human cDNA. Relative values were calculated by normalizing to the values obtained with the original protocol (Drop-seq), the error is shown as standard deviation. P-values were determined on the normalized values with a t-test.

As the two proposed on-chip de-emulsification processes present substantial changes to the original approach, we decided to determine their impact on the molecular integrity of the samples. In particular, since wash buffers that potentially contain environmental RNAs are flown over the bead sediment layer of the cp-chip, we focused on quantifying cross-contamination of the new approach. To this end, species-mixing experiments are conventionally used. However, turnover times for high- throughput sequencing and data-analysis are still not time nor cost-effective, making species-mixing experiments impractical for protocol optimization. Hence, to conveniently measure the impact of bead capture using the cp-chip on the existing experimental setup, we developed an scRNA-seq species purity test based on quantitative PCR (Figure 2.4A). Specifically, we hybridized purified human mRNA derived from HEK 293T cells on oligonucleotide capture beads, and encapsulated them into microdroplets. In parallel, we generated a microemulsion carrying free-floating *Drosophila* melanogaster RNA in its aqueous phase. Next, both emulsions were mixed, simulating a ‘worst-case scenario’ sample made of few droplets containing beads and a vast number of bead- less droplets containing contaminant mRNA. This way, the performance of different bead capture approaches can be easily quantified through STAMP formation followed by qPCR targeting human and Drosophila genes. The relative ratio between quantified human and Drosophila DNA thereby reflects the signal-to-noise ratio of the purification protocol. As target genes for quantification, we chose Rp49 and GAPDH, two abundant housekeeping genes of *Drosophila* and humans, respectively. By hybridizing equal amounts of *Drosophila* and human RNA to beads, we observed sufficient specificity (data not shown) and similar amplification characteristics for both target genes (Figure 2.4B).

Applying this benchmarking approach to the previously developed purification procedures, we found that, in comparison to the original approach, bead capture from broken emulsions showed less cross-contamination when a large number of beads (up to 10 000 beads) was used (Figure 2.4C). As it is unlikely that the microfluidic chip itself reduced cross-contamination, we assumed that one additional wash step prior to bead capture led to this result. To validate this hypothesis, we added an additional wash step to the original Drop-seq protocol, which yielded a similar decrease in back-ground as observed for the cp-chip-based bead purification (Supplementary Figure 2.1). In contrast, direct breakage of the emulsion on-chip for bead recovery significantly increased species cross-contamination by approximately 2.5-fold when a large number of beads was used (Figure 2.4C). These results suggest that direct emulsion breakage is to be avoided when handling a large number of beads, but could still be useful for small samples that are challenging to process. We hypothesized that as soon as beads are occupying the whole pillar array, beadless contaminating droplets are forced into contact with beads leading to cross-contamination. To validate this hypothesis, we performed the same experiments with lower bead numbers (up to 1000 beads). Interestingly, we found that the background from contaminating droplets was significantly reduced for small bead numbers compared to large bead numbers for direct emulsion breakage, indicating that this approach is indeed viable for small samples (Figure 2.4C).

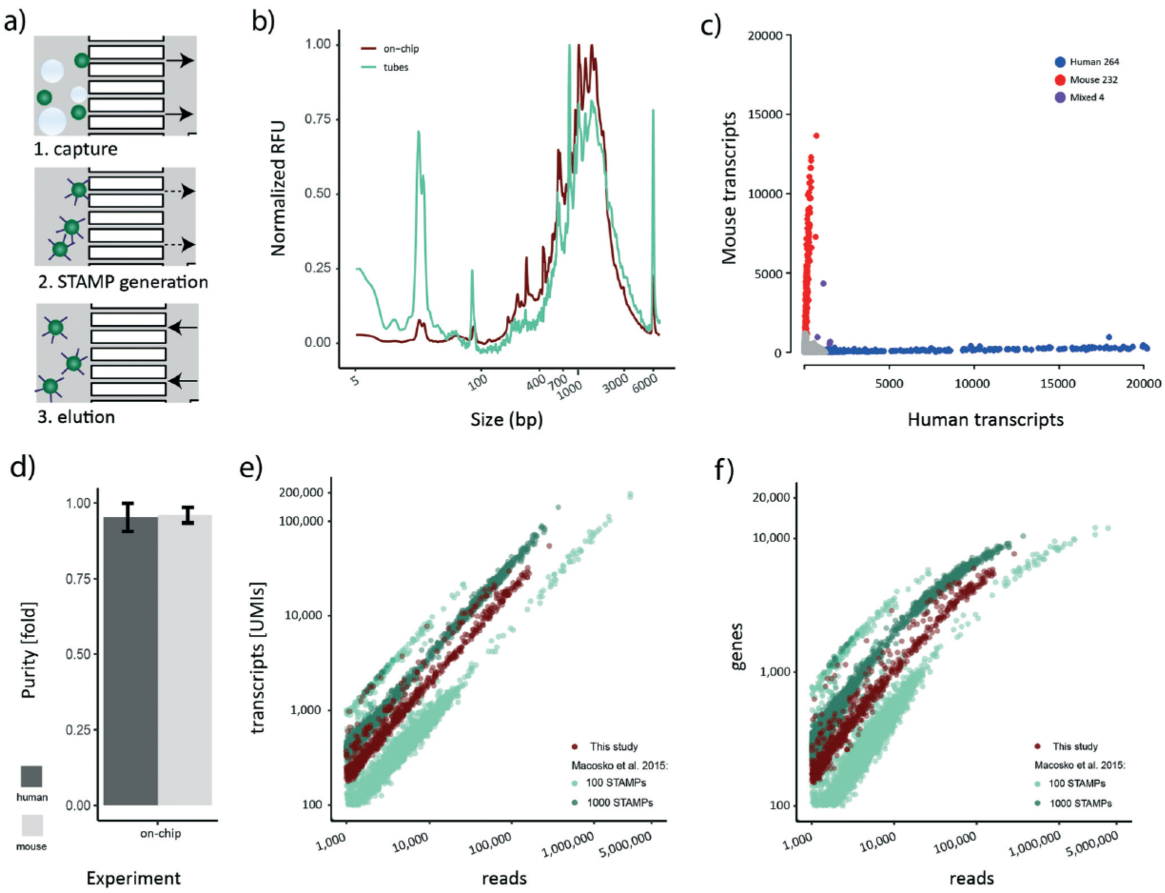


Figure 2.5 - STAMP generation on-chip: (A) for on-chip STAMP generation, beads from a broken emulsion were captured on the cp-chip (arrows indicate flow direction), and reagents for reverse transcription and exonuclease treatment injected onto the chip. After all reactions and washes, beads were eluted by reversing flow. PCR amplification was performed in a tube. **(B)** cDNA profile obtained after STAMP generation in tubes and on-chip. Relative fluorescence units (RFUs) of each sample were normalized to the maximum value of the respective sample (1037 RFU for the on-chip sample and 122 RFU for the tube sample) in the range of 2 bp–21 kb. **(C)** Transcript numbers per cell for mouse and human transcripts. Cells classified as belonging to one species (>90% species purity). Mouse cells are colored red, human cells blue. Purple points represent cells containing mixed transcripts (<90% species purity). **(D)** Average single cell purity for mouse and human cells. **(E)** Number of UMIs detected per read for HEK293T cells in the Macosko *et al.* 2015 100 STAMP and 1000 STAMP species mixing experiment, and for this study. **(F)** Number of genes detected per read for HEK293T cells in the Macosko *et al.* 2015 100 STAMP and 1000 STAMP species mixing experiment, and for this study.

Although bead losses are not as pronounced for the STAMP generation process compared to de-emulsification (Figure 2.3A), the ten manual handling steps that are required are both error-prone and labor-intensive. To address this issue, we next aimed to explore the possibility of implementing the bead processing steps until the final PCR step on the cp-chip itself, minimizing bead loss and manual work for the whole protocol. Specifically, we aimed to mimic a routine experimental size of approximately 1000 STAMPs, captured from broken emulsions. To test this approach, we first encapsulated a mix of HEK 293T cells and murine brown preadipocytes at 100 cells per µL on the e- chip, broke the emulsion in vessels and purified the beads (~10 000) on the cp-chip. Subsequently, we performed RT and ExoI treatment directly on-chip. Finally, the beads were eluted from the cp-chip, bead-bound cDNA was PCR-amplified in a reaction tube, and the cDNA profile analyzed (Figure 2.5A). As a control, an emulsion containing the same amounts of beads and cells was generated, beads captured, and STAMP generation performed in tubes according to the original Drop-seq protocol. During on-chip STAMP generation, beads were clustering loosely, yet no clumping was noted. We observed that on-chip STAMP generation yielded high-quality cDNA traces compared to cDNA from STAMPs generated in tubes (Figure 2.5B), indicating that it does not negatively impact polyA RNA integrity. The most striking difference between both traces was a strongly pronounced peak in STAMPs produced in tubes at approximately 25 bp, which most likely reflects residual primer. As subsequent bead purification will remove small fragments, this putative primer presence can be neglected.

Our novel bead processing approach introduces a variety of changes in the overall workflow: new sample vessels, encapsulation chip, bead recovery strategy, and STAMP generation process. Potentially, this could impair the scRNA-seq data quality at multiple levels, including decreased single-cell accuracy (single-cell purity) and altered sensitivity. To address these concerns, we decided to sequence the previously pre- pared species-mixed library that was generated utilizing a) the e-chip, b) efficient bead recovery from a broken emulsion using the cp-chip, c) the on-chip STAMP generation protocol, and d) PCR amplification of eluted STAMPs in tubes. We approximated the number of STAMPs in this experiment to be around 500. The resulting Barnyard plot showed a clear organism-specific transcriptome separation (Figure 2.5C). To benchmark the single-cell quality, we analyzed mean species purity, which we determined to be 96% for mouse cells and 95% for human cells (Figure 2.5D), well within range of the original protocol at 100 cells per µL[32].

Finally, we set out to determine whether the new workflow offered a similar data quality as the original published datasets[32]. Specifically, we processed the 100 STAMPs and 1000 STAMPs species mixing datasets from the original publication and merged them to account for inter-experimental deviations. For downstream analysis, only HEK 293T cells were maintained since they represent the common cell line for both studies. As each dataset contains a different number of STAMPs and is sequenced to varying read depth, we com- pared the amount of UMIs detected to the number of sequencing reads obtained per cell (Figure 2.5E). Reassuringly, all datasets showed a comparable linear relationship between the number of UMIs detected per obtained sequencing read. This linearity suggests a similar transcript diversity across all samples, and thus a comparable sensitivity. Importantly, despite various experimental differences such as different bead- batches etc., we observed that the on-chip processed dataset clusters in-between the experimental data derived from the original data, suggesting comparable performance for both protocols. Additionally, we analyzed the gene numbers in relation to the obtained sequencing reads (Figure 2.5F). Similar to the transcript counts, we observed that a comparable number of genes per read is detected in both datasets. Furthermore, gene numbers for all datasets were saturating at around 10 000 genes per cell, further confirming the comparable sensitivity among all approaches. Overall, these findings show that our developed microfluidic bead processing workflow achieves a similar experimental quality as the original protocol, but with the advantage of constituting a simpler, more flexible, and more efficient approach.

2.4 Discussion

In this study, we present the development of an optimized Drop-seq workflow that simplifies the experimental setup, enhances the overall bead recovery efficiency by more than two- fold and significantly simplifies the bead processing protocol. This workflow involves a newly developed encapsulation chip that is compatible with both syringe-pump or pressure-based driver setups. Utilizing an air pressure-based setup, the expensive magnetic stirrer was replaced by inexpensive laboratory equipment. Consistent with previous results, we found that the original bead processing strategy, especially during de-emulsification, is prone to large sample losses. Here, we were able to overcome this limitation by developing a dedicated bead capture and processing chip, which enables high-efficiency bead capture and STAMP generation. Using this cp-chip, we explored its potential application to varying experimental situations. In the first operation mode, developed for larger samples, the device was used to capture beads from broken emulsions for efficient bead capture and convenient STAMP generation. We found that bead capture from broken emulsions on the cp-chip decreased cross-contamination com- pared to the original Drop-seq protocol, which we attributed to the inclusion of an additional washing step in the new protocol rather than to a chip-related property. Beyond the cp-chip-based bead processing approaches, this protocol modification has the potential to increase single-cell purity in high cell-density samples using the original protocol. For the second operation mode, we explored the possibility of directly capturing beads from droplets on-chip, which we found to be the most efficient strategy for de-emulsification. For this approach, we observed a cross-contamination increase of 2.5-fold when introducing a large number of beads. We hypothesized that full bead occupation of the cp-chip pillar array forced beads in contact with contaminating droplets. Indeed, we found that direct emulsion breakage on-chip using smaller bead numbers showed cross-contamination levels be- low the ones observed for the original Drop-seq protocol, thus representing a viable processing strategy for rare or low- input samples. Finally, we showed that STAMP generation can be efficiently performed on-chip, eliminating all manual handling steps post-bead capture. Complementary to this approach, we established an inexpensive and rapid qPCR method to explore suitable ways to integrate the bead processing approach in an existing experimental workflow. Finally, we showed that our new workflow features a comparable single-cell purity and sensitivity to the original publication.

Globally, our new workflow adds important features for routine experimentation, especially on medium to small samples. Furthermore, although small samples are still rarely used with Drop-seq, future developments of high cell capture efficiency protocols such as the one presented by Chung and colleagues[108] will make efficient bead processing strictly necessary. Thus, we expect that our microfluidic-based bead processing will constitute an important experimental cornerstone of these next generation technologies.

2.5 Material and methods

2.5.1 Microfluidic chip design and flow simulations

Chips were designed using Tanner L-Edit CAD software version v 2016.2 (Mentor). Flow simulations were carried out using COMSOL Multiphysics version 5.2. Segments were extruded to the same height as the actual device (80 µm), flow conditions were set to “creeping flow”, incompressible flow, water density with no-slip boundary conditions. The inlet and outlet were separated by a pressure gradient of 5 PSI. The chosen mesh type was free triangular with size option set to normal. The velocity profile of the slice running at half channel height was plotted.

2.5.2 Soft-lithography and microfluidic device fabrication

6-Inch chromium masks were exposed in a VPG200 laser writer (Heidelberg instruments) using a 20 mm laser writing head. Masks were developed using an HMR 900 mask processor (Hamatech). 60 µm (e-chip) or 80 µm (cp-chip) thick SU8 photoresist layers were deposited with an LSM-200 spin coater (Sawatec) and then exposed on a MJB4 single side mask aligner (SussMicroTec) and manually developed. The SU8 processing steps were done according to manufacturer's instructions for the 3050 series (Y311075 0500L 1GL, Microchem). Developed wafers were used as a mold for PDMS chips after passivation (10 : 1 ratio PDMS : curing agent). PDMS and curing agent were mixed, degassed, and poured on the mold. The PDMS mixture was cured for 1 hour at 80 °C. Next, inlet holes were punched, and the chip surface activated with oxygen plasma (45 s at ~500 mTorr). The chips were bonded on surface activated glass slides and incubated at 80 °C for at least 2 hours. Before use, the chips (e-chip) were surface-treated with 2% Trichloro (1H,1H,2H,2H-perfluorooctyl) silane (448931-10G, Sigma-Aldrich) dissolved in 3M HFE 7500 (297730-93-9) for 5 min.

2.5.3 Procedures for microfluidic device handling

The MFCS-EZ (2 × 345 mbar, 1 × 1045 mbar outlet version) air-pressure driver (Fluigent) was connected to the sample vessels (CG-4909-04, Chemglass) using silicone tubing (1175- 8705, Fisher Scientific) and blunt-end needles (23G). Vessels were connected to the microfluidic chip using Tygon tubing (06420-02, Cole-Parmer) as shown in Figure 2.2E. The vessels were loaded with cells, beads and EvaGreen droplet generation oil (186-4006, Bio-rad). To maintain constant bead mixing, the plastic adaptor for a vortex shaker (Figure 2.2D) was designed in Fusion 360 (Autodesk) and 3D printed in ABS using a M200 3D-printer (Zortrax). Pressure for the bead and cell vessels was set to 345 mbar and for oil phase pressure to 900 mbar. The tubing was directly connected to the microfluidic chip placed on an inverted microscope. The vessels were pressurized in the following order: 1. cell suspension, 2. bead suspension, and 3. oil phase. After droplet formation stabilized and residual polydisperse emulsions exited the collection tubing, droplets were collected in a 50 mL tube. Droplet formation was monitored using a HotShot 1280CC (NAC) high-speed camera.

2.5.4 On-chip droplet breakage and bead recovery strategy

After droplet collection, droplet and oil phase were loaded into a 200 µL pipette tip (2239915, Bio-Rad) and directly injected into the cp-chip. Once beads were recovered on the pillars, washing was performed twice with 6× SSC buffer. To elute the beads from the chip, Tygon tubing was connected to the inlet, the end placed into a microcentrifuge tube, and 6× SSC injected into the outlet.

2.5.5 Off-chip droplet breakage and bead recovery strategy

After droplet collection, the oil phase was removed from the tube and 30 mL 6× SSC buffer was added to the emulsion in the 50 mL tube. Next, 0.5 mL 1H,1H,2H,2H-Perfluoro-1- octanol (370533-5G, Sigma-Aldrich) was added, and the tube was agitated by vigorous shaking. Following this, two washing steps were performed with 6× SSC at 1000×g by lowering the brake by 50% without disturbing the bead interphase. Subsequently, as much 6× SSC as possible was removed with- out disturbing the beads, leaving only 5 mL of oil–water phase. Both 6× SSC and perfluoro-1-octanol phases were transferred to the bead recovery chip, as described above. After bead capture, the beads were washed two times with 6× SSC to remove any residual oil and perfluoro-1-octanol. Beads were recovered in a microcentrifuge tube by injection of 6× SSC buffer into the chip outlet (as described above).

2.5.6 Original Drop-seq droplet breakage and bead recovery strategy

Bead recovery from droplets was performed as described in the latest published online version of the protocol.25 For experimental results shown in Supplementary Figure 2.1, an additional washing step was included using 6× SSC prior to pelleting the sample in Falcon tubes followed by the transfer into Eppendorf tubes.

2.5.7 Bead quantification

To evaluate bead recovery strategies, bead numbers were quantified prior to the experiment, and after all processing steps. The bead starting amount was enumerated by scanning the whole sample on a V600 scanner (Epson). Specifically, all beads were pipetted into a 48 well plate lid, placed on the scanner. After a ~3 min sedimentation period, the lid was scanned at 4800 dpi resolution in transparency mode. Next, the beads were carefully retrieved from the plate lid and placed back into a microcentrifuge tube. An additional scan of the empty lid was used to quantify the remaining beads, which were subtracted from the starting amount. After completion of the experiment, quantification of the remaining beads was executed as above. The ‘sampling loss’ amounted on average to 3% of the processed beads, for each manual transfer of bead solutions between enumeration plate and experimental device. The obtained images were quantified using Fiji (ImageJ version 1.50g) by thresholding, watershed separation, and particle detection based on a size and circularity threshold. Systematic errors of the quantification (e.g. closely touching beads) were manually corrected in all images.

2.5.8 Cell handling

For species mixing experiments, HEK-293T (ATCC Cat. No. SD-3515) and murine brown preadipocyte (provided by Prof Christian Wolfrum's laboratory, ETH Zürich)[109] cell lines were used. Cells were cultured up to 90% confluency in Glutamax DMEM (61965026, Thermo Fisher Scientific) supplemented with 10% FBS (10270-106, Thermo Fisher Scientific) and Pen- Strep (15140-122, Thermo Fisher Scientific). Following this, cells were dissociated using Trypsin-EDTA (59418C, Sigma-Aldrich), washed once with PBS (14040091, Thermo Fisher Scientific) and counted using Trypan-blue live-dead stain (T10282, Thermo Fisher Scientific) using a Countess (Invitrogen) cell counter. Cells were finally re-suspended in PBS supplemented with 0.01% BSA and murine RNAse inhibitor (M0314S, NEB), and mixed in a 1 : 1 ratio to an adjusted concentration of 100 cells per µL.

2.5.8 qPCR assay for performance test of optimized purification strategies

Total RNA from *D. melanogaster* tissue and HEK-293T (human) cell line was isolated using the Direct-zol RNA miniprep kit (R2056, Zymo Research). Beads were incubated with the extracted human RNA at 110 ng µL−1 for 10 min at room temperature with mixing. Unbound RNA was washed away and beads with bound RNA were encapsulated in droplets. Following this, droplets containing free-floating *D. melanogaster* RNA at 110 ng/µL concentration were collected. Both groups of droplets were mixed in a 1 : 1 ratio and after that, droplet breakage and bead recovery for reverse transcription were performed. Once STAMPS were generated, cDNA was amplified and purified from sub-sampled 800 beads following the original protocol. The purified cDNA was used for qPCR analysis of species cross-contamination. 0.75 ng of cDNA was amplified in PowerUp SYBR Green MM (Thermo Fisher Scientific) containing 200 nM of either human specific GAPDH primer mix (forward: acccactcctccacctttgac, reverse: tgttgctgtagccaaattcgtt) or *D. melanogaster* specific Rp49 primer mix (forward: gacgcttcaagggacagtatctg, rev: aaacgcggttctgcatgag). Amplification was done in technical triplicates on a StepOnePlus RT-PCR System (Applied Biosystems). As a control, all three droplet breakage strategies were compared to beads bound to HEK-293T RNA only, *D. melanogaster* RNA only and combined HEK-293T - *D. melanogaster* (1 : 1 ratio) RNA to ensure primer specificity. Ct values were thresholded at 0.1 RFU and the relative ratio between HEK-293T and contaminating *D. melanogaster* cDNA was used to assess the cross-contamination/background of each purification protocol.

2.5.9 Library preparation and sequencing

Following droplet breakage and bead (SeqB lot 120817, ChemGenes) recovery, reverse transcription (RT), exonuclease I (ExoI) treatment and PCR were performed as described in the original protocol[32]. Libraries were purified using Ampure XP beads (0.6× ratio to remove small fragments), cDNA was quantified using a Qubit HS kit (Thermo Fisher Scientific) and integrity analyzed on a Fragment Analyzer (Agilent). Libraries were prepared using in-house produced Tn5 loaded with adapters, as described.27 Size selected and purified libraries were sequenced paired-end on a NextSeq 500 system (Illumina) in High-Output mode following recommendations from the original protocol (read 1 20 bp and read 2 50 bp).

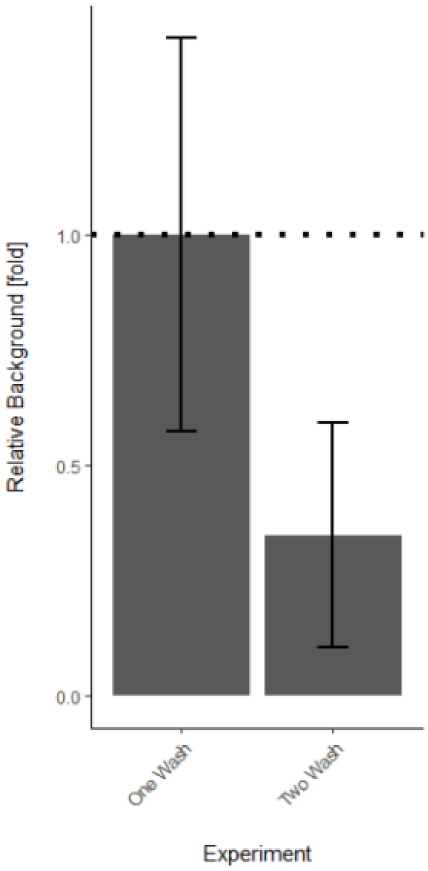
2.5.10 STAMP generation on chip

Beads were recovered from broken emulsions on-chip with- out any additional incubation times. Subsequently, RT, ExoI treatment and all washing steps were performed on-chip. To compensate for liquid evaporation from the PDMS device during RT and ExoI treatments, the reaction mix was injected every 15 min into the device allowing complete bead immersion. The microfluidic chip containing the beads was incubated at temperatures and durations as indicated in the original protocol[32] using a flat thermoblock. After the last washing steps, beads were eluted in an Eppendorf tube for counting and PCR amplification. Library preparation was performed as described above.

2.5.11 Data analysis

The data analysis was performed using the Drop-seq tools package[32] on the Vital-IT HPC platform. After trimming and sequence tagging, reads were aligned to the mixed human : mouse reference genome (hg38 and mm10) using STAR (version 2.5.3.a)[110]. Following the alignment, the gene annotation was added, bead synthesis errors were corrected and cell barcodes extracted. Subsequently, the BAM files containing the processed data were filtered, split into mouse and human annotated BAM files, and digital gene expression matrices were generated for each species. Preliminary data-analysis was done in ASAP[111]. Downstream data analysis was done using R (version 3.5.1), plots generated using the R package ggplot2 (version 3.0.0).

2.5 Supplementary figures



Supplementary Figure 2.1 - Impact of bead washing on background The figure shows the impact of adding one more washing step (introduced in the bead capture protocol when utilizing the cp-chip) compared to the original protocol. One emulsion comprised of droplets containing beads hybridized to human RNA, and droplets containing contaminating Drosophila RNA was generated. The emulsion was split, with one half being processed using the original Drop-seq protocol (One Wash), and the other half using the same protocol but adding one more washing step (Two Wash).

Chapter 3: 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], Vincent Gardeux[1,2], 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

Shared First Author Contributions: MB and JB conceptualized the study, JB developed the microfluidic chip, JB developed the machine-vision system, MB and JB performed all benchmarking experiments, MB performed library preparation, MB analyzed benchmarking experiments, MB and JB performed all organoid experiments, JP, JB and MB analyzed the scRNA-seq data, JB and MB wrote the manuscript.

Chapter information: This chapter is published on Biorxiv
https://doi.org/10.1101/2020.05.19.103812

3.1 Abstract

Single-cell RNA-sequencing (scRNA-seq) has transformed our ability to resolve cellular properties across systems. However, current scRNA-seq platforms are one-size-fits-all approaches that are tailored toward large cell inputs (> 1,000 cells), rendering them inefficient and costly when processing small, individual tissue samples. This important drawback tends to be resolved by loading bulk samples, but this yields confounded mosaic cell population read-outs. To overcome these technological limitations, we developed a deterministic, mRNA-capture bead and cell co-encapsulation dropleting system, DisCo. We demonstrate that DisCo enables precise particle and cell positioning and droplet sorting control through combined machine-vision and multilayer microfluidics. In comparison to other microfluidics systems, the active flow control driving DisCo, enables continuous operation and processing of low-input samples (< 100 cells) at high capture efficiency (> 70%). To underscore the unique capabilities of our approach, we analyzed intestinal organoid development by "DisCo-ing" 31 individual organoids at varying developmental stages. This revealed extensive organoid heterogeneity, identifying distinct subtypes including a regenerative fetal-like *Ly6a*$^+$ stem cell population which persists as symmetrical cysts even under differentiation conditions. Furthermore, we uncovered a so far uncharacterized "gobloid" subtype consisting predominantly of precursor and mature (*Muc*2$^+$) goblet cells. These findings demonstrate the unique power of DisCo in providing high-resolution snapshots of cellular heterogeneity among small, individual tissues.

3.2. Introduction

Single-cell RNA sequencing (scRNA-seq)[112] 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 [32,37,103,113,114], have led to a substantial increase in experimental throughput, enabling massive mapping efforts such as the mouse and human cell-atlas studies[96,114,115]. These developments were accompanied by biochemical advances, for instance for targeted transcript detection or library multiplexing[116,117], 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. 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[59,118]. 3) Reduced effectiveness at low inputs because of limited cell capture efficiencies or cell size-selective biases[39] when processing small heterogeneous samples. To illustrate these limitations, we summarized the performance of various scRNA-seq technologies on low input samples in Supplementary Table 3.9. Consequently, small samples, involving for instance zebrafish embryos[119], organisms like *C. elegans*[45], or intestinal organoids[120–122], 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, making this platform well suited to handle small, individual tissues. Here, we exploit DisCo's unique capabilities to explore the heterogeneous early 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[120]. These unpredictable developmental patterns represent one of the major limitations of this model system, preventing their widespread implementation e.g. in drug screens[123]. 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[124,125], 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[120]. 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*)[126–129] and that persist under differentiation conditions. In addition, we uncovered a rare subtype that is predominantly comprised of precursor- and mature goblet cells, which we term "gobloids".

3.3 Results

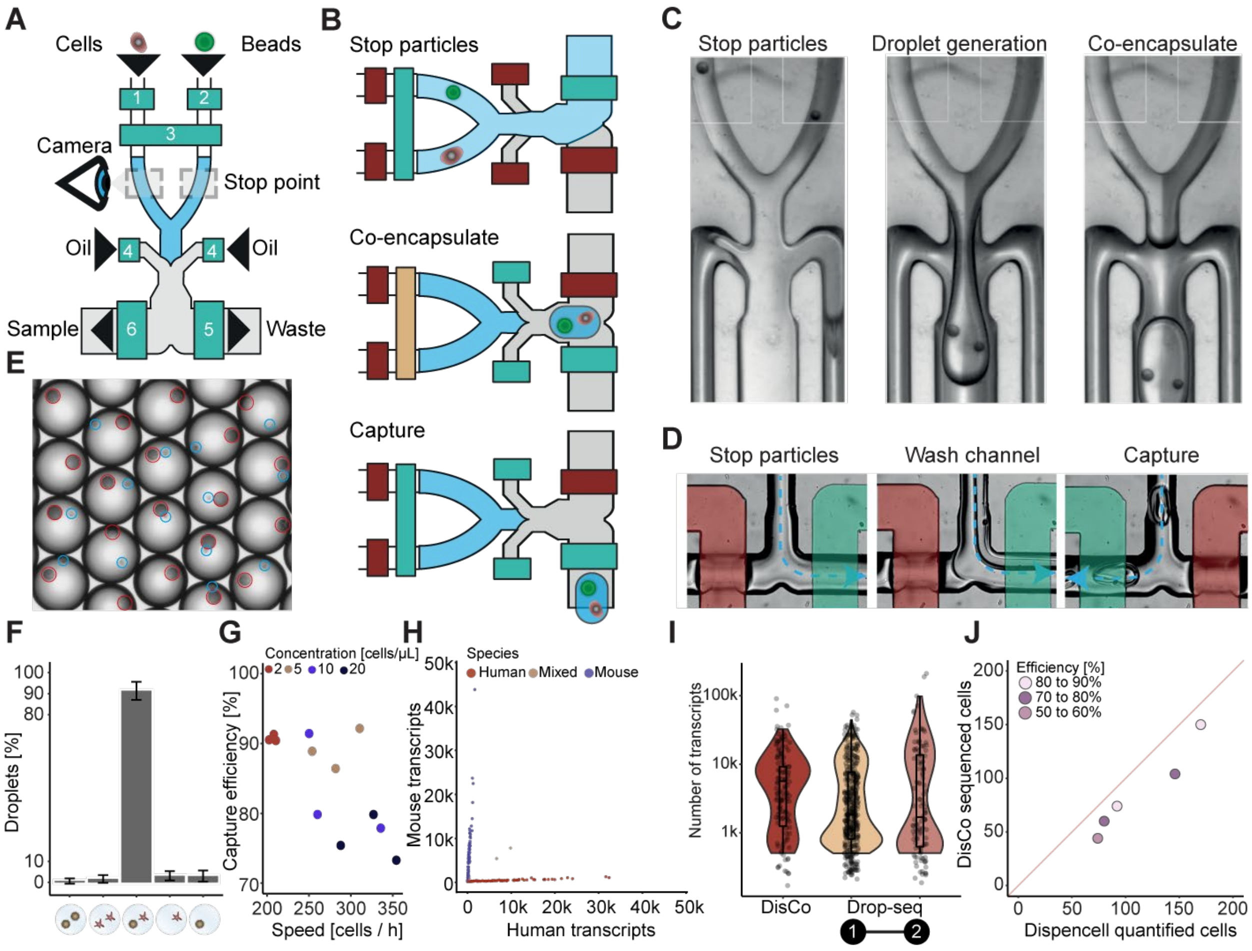


Figure 3.1 - Overview and critical feature assessment of the deterministic co-encapsulation (DisCo) system: **(A)** Schematics of the DisCo microfluidic device. The device contains three inlet channels for cells, beads, and oil, and two outlets for waste and sample liquids. All inlets and outlets are 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 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 standard deviation. **(G)** Cell capture efficiency and cell capture 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 Bičanin, Bues *et al.* 2019[130] and [2] - from Macosko *et al.* 2015[32]), 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 recovery efficiency after sequencing.

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 3.1A, full design Supplementary Figure 3.1A). On this device, each inlet and outlet was augmented with a Quake-style microvalve[131], 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 3.1B): 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 3.1C). To enable precise coordination of particles in microchannels, we developed a machine-vision-based approach utilizing subsequent image subtraction for blob detection (Supplementary Figure 3.1B), 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 3.1C), 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 3.1D). 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 3.1E). Post droplet formation, the outlet valves were actuated to separate the formed droplet from the excess waste liquids (Figure 3.1D). With all components operating in tight orchestration, we were able to generate monodisperse emulsions with high co-encapsulation purity (Figure 3.1E).

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[32] approach as it only requires coordination of two channels, as compared to three channels for inDrop[37]. 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 3.1F). Overall, the system provided high cell capture efficiencies of 90% at around 200 cells per hour for a 2 cells/µL cell concentration (Figure 3.1G). 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[130]. 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 3.1H), consistent with the limited number of previously detected doublets (Figure 3.1F), and increased read-utilization rate compared to conventional Drop-seq experiments (Supplementary Figure 3.1F). As previously reported[132], we found that our data displayed a skewed barcode sequence editing distance distribution compared to a true random distribution (Supplementary Figure 3.1G). 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 (as described in Material and Methods). We found that this approach did not compromise the single cell purity (Supplementary Figure 3.1H) and improved the detectable number of transcripts per cell as compared to published Drop-seq datasets on HEK 293T cells[32,130] (Figure 3.1I).

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 3.1I). 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 successfully co-encapsulated, which corresponds to a co-encapsulation efficiency of 91.6% (SD ± 1.6%) of all visible cells, while 74.9% (SD ± 10.7%) of input cells were found as barcodes over 500 UMIs per cell (Figure 3.1J).

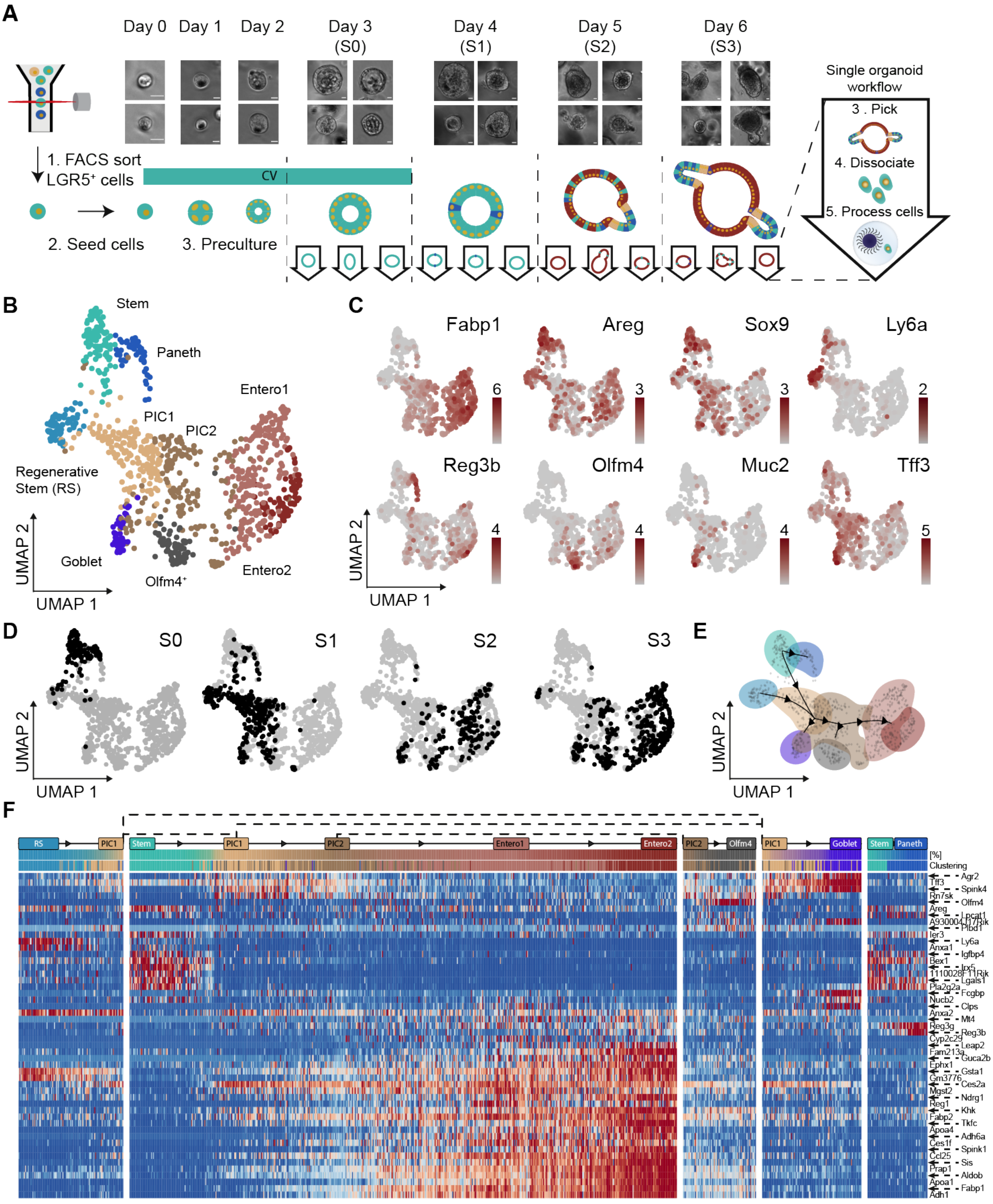


Figure 3.2 - Utilizing DisCo to map intestinal organoid cell heterogeneity along development: (A) Overview of the experimental design for DisCo'ing individual organoids. Single LGR5⁺ intestinal stem cells were isolated via FACS and precultured for 3 days under stem cell maintenance conditions (ENR CV 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 50 µm. **(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 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[133]. 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)[134], Enterocytes cluster 1/2 (Entero1/2).

As a real-world application, we used DisCo to explore the developmental heterogeneity of intestinal organoids[135]. 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)[135]. When grown from single stem cells, organoids of very different morphologies form under seemingly identical *in vitro* conditions (Figure 3.2A, overview image in Supplementary Figure 3.2A). Pooled tissue scRNA-seq data has shed light on the *in vivo*-like cell type composition of these organoids[120–122,136], 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 responsible for crypt formation[120]. 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)[137]. 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 3.2A). 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 encapsulated cells was noted and correlated to the number of barcodes retrieved, which was in approximate accordance (Supplementary Figure 3.2B). The even distribution of the number of reads mapping to ribosomal protein transcripts 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 3.2C).

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[121,136] since it revealed expected cell types including *Fabp1*-expressing enterocytes, *Muc2*-expressing goblet cells, *Reg3b*-positive Paneth cells, and *Olfm4*-expressing stem cells (Figure 3.2B and 3.2C). 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 3.2D). Noteworthy, we found that batch effects are correctable since no batch-based clustering was observed after correction (Supplementary Figure 3.2E). We also did not detect any clustering driven by cell quality, e.g. detected transcripts or mitochondrial transcripts (Supplementary Figure 3.2C). These findings support the cell type-resolving power of our DisCo platform (Figure 3.2C, extensive heatmap in Supplementary Figure 3.2F). 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 3.2D), and increased YAP-1 target gene expression (Supplementary Figure 3.2G), suggesting that these cells are most likely regenerative fetal-like stem cells[128,129,138]. 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)[134], given their occurrence at early developmental time points (Figure 3.2D). 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[133] to infer lineage relationships between cell types and to identify genes that may be of particular significance for waypoints along differentiation (Figure 3.2E). Beyond the previously utilized marker genes for cell type annotation, for example *Reg3b* and *Reg3g* for Paneth cells, additional established markers[139] were identified, such as *Agr2 and Spink4*, and *Fcgbp* for goblet cells (Figure 3.2F). Overall, this suggests that the meta-data produced with our DisCo platform aligns with and expands prior knowledge.

Intriguingly, we observed 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[127], 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 3.3A). 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 comprised 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 3.3A). To confirm the presence of *Ly6a*$^+$ organoids in our cultures, we utilized RNAscope (Figure 3.3B, controls Supplementary Figure 3.3B) 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.

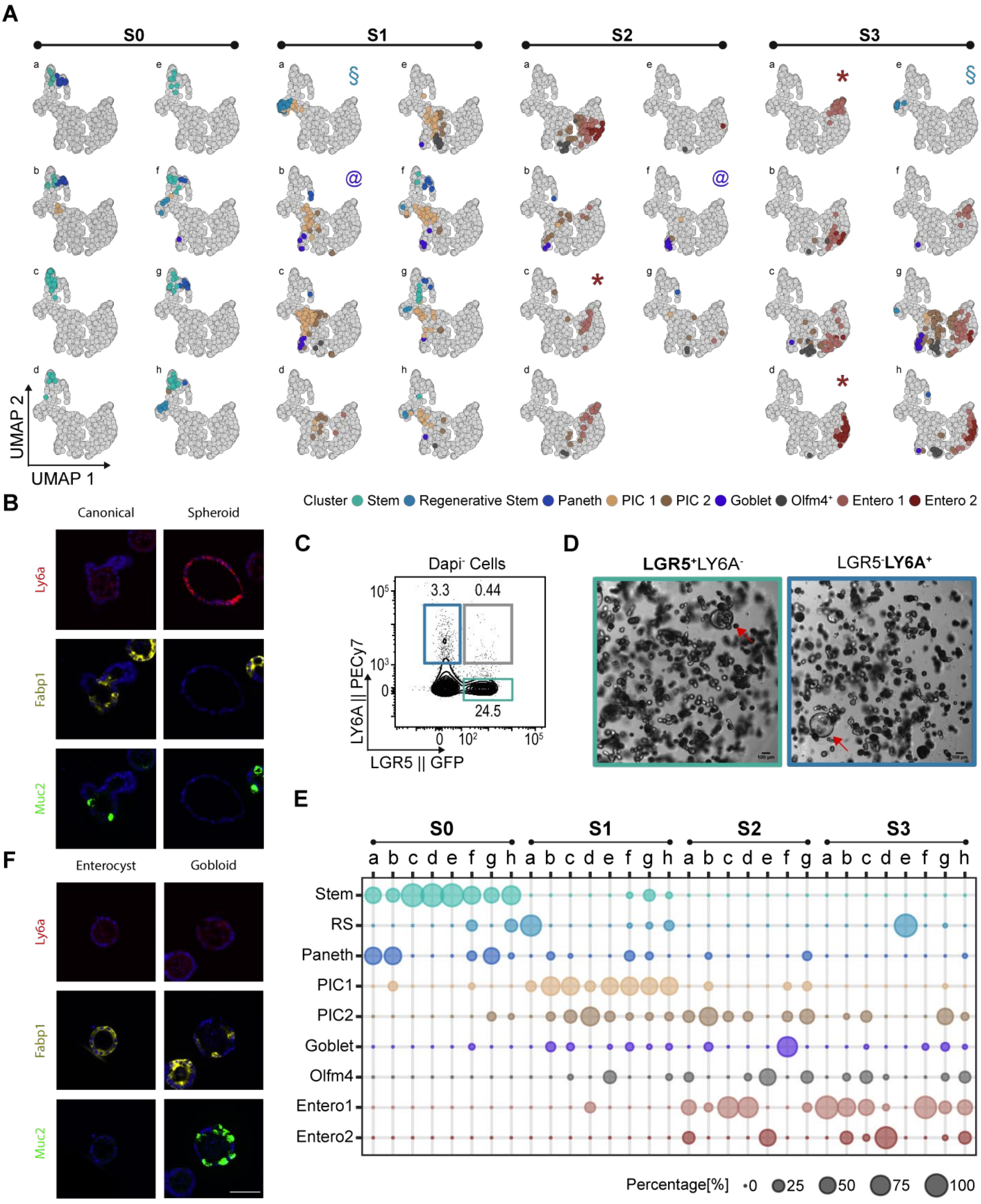


Figure 3.3 - Cell type distribution and marker gene expression across individual intestinal organoids during development: (A) Projection of cell types onto 31 individual organoids. Cells per single organoid were colored according to their global clustering and highlighted on the UMAP embedding of all sequenced cells. Projections are grouped according to their sampling time. Manually classified organoids were annotated with the following symbols: "*" enterocysts, "§" spheroids, "@" gobloids. **(B)** *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 3.2a 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.

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 3.3C). This finding, in combination with our trajectory analysis (Figure 3.2E and 3.2F), 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 3.3D). These results indicate that LGR5$^-$ LY6A$^+$ cells have full stem cell potential, comparable to that of previously described fetal-like stem cells[127]. Furthermore, the fact that LGR5$^-$ LY6A$^+$ cells did not display a propensity towards spheroid formation suggests that environmental conditions, e.g. matrix stiffness, rather than the initial cell state dictate the formation of spheroids.

Beside the *Ly6a*$^+$ cell-enriched organoids, our data suggested the presence of additional organoid subtypes in the per organoid mappings (Figure 3.3A). 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 3.3E), and marker gene expression in individual organoids (Supplementary Figure 3.3C). 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 3.3A). To detect more subtle molecular differences, we used psupertime[140] 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 3.3D).

While organoids dominated by enterocytes were previously described as enterocysts[120], 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 3.3F, controls in Supplementary Figure 3.3B). In addition, and 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”.

3.4 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[32,37], 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 (Supplementary Table 3.9). 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[141,142]. In concept, DisCo is thus fundamentally different to passive particle pairing approaches such as traps[108,143,144] 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[39,145]. This renders the DisCo approach universally applicable to any particle co-encapsulation application[146,147], 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 *Drosophila* larvae[46] and, as shown here, single intestinal organoids. 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[148,149]. Indeed, substantial cell loss was a regular occurrence, even with optimized dissociation and processing strategies (see Materials and 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[120,121,136]. 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[128,129,138]. This subtype, previously described under alternate culture conditions as spheroid-type organoids[122,126,127], 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. Of particular interest was one organoid subtype that we termed "gobloid" given that it predominantly comprises 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[150], may drive gobloid development[151]. However, failure to produce Paneth cells has previously been suggested as a mechanism underlying enterocyst development[120], 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 organoids is a powerful approach to explore tissue heterogeneity and to yield new insights into how this heterogeneity arises. In comparison to established approaches such as automated microscopy[120,122], DisCo is magnitudes lower in experimental scale. Nevertheless, scRNA-seq data acquired from 31 organoids enabled us to recapitulate previous findings, benchmarking DisCo, and most importantly, to uncover novel subtypes, 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 expect this approach to be applicable to rare, small clinical samples to gain detailed insights into disease-related cellular heterogeneity and dynamics.

3.5 Acknowledgements

We thank Wanze Chen and Petra C. Schwallie 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.

3.7 Materials and Methods

3.7.1 System comparison metrics

Performance metrics for (Supplementary Table 3.9) were calculated the following ways:

- Minimum cell input estimates: The minimum cell input values were derived from the following sources: 10X Chromium[43]: Lowest cell input number from the 10X Chromium manual (CG000183 Rev C); inDrop[37]: Lowest numbers mentioned in the 1CellBio manual (Single Cell Encapsulation Protocol, Version 2.4); Drop-seq[32]: 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; FACS based methods[152,153]: Input limits as described by Hwang *et al.* 2018[59]; Fluidigm C1[145]: Lowest cell input number from the Fluidigm C1 specification sheet (Specification Sheet PN 101-3387 D1); Wafergen iCell8[154]: 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[101]: The lowest cell number used for capture in Gierahn et al. 2017[101]. 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, inDrop, and Drop-seq were derived from Zhang *et al.* 2019[44] 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). The efficiency for the Fluidigm C1 system was derived from Xin *et al.* 2016[40] from a high input sample of primary cells. The efficiency represents a conversion efficiency from captured to sequenced cells passing quality thresholds, thus it does not include cell capture inefficiencies which are substantial at low cell inputs (personal communication Dr. Bastien Mangeat, Gene Expression Core Facility EPFL). For the Wafergen iCell8 system, an efficiency estimate was derived from Wang et al. 2019[50] 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[101] 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)[149]. Run costs for 10X Chromium, Fluidigm C1 (96), and Wafergen iCell8 were derived from Wang *et al.* 2019 (Table 2)[50]. 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).

3.7.2 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.

3.7.3 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 3.1B. 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 3.1C). 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.

3.7.4 Microfluidic chip design and fabrication

The design of the microfluidic chip for deterministic co-encapsulation is presented in Supplementary Figure 3.1A. 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°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.

3.7.5 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[14] 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.

3.7.6 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[130] (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°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°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[15] for 6 minutes at 55°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)[16]. Material and reagents are listed in the Material and reagent list, points 3 - 10.

3.7.7 Mammalian cell culture handling for 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.

3.7.8 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 3.1F).

3.7.9 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.

3.7.10 Organoid cell culture and handling

Isolation of the Lgr5-eGFP^{+} stem cells and initial culture was performed as previously described[17]. 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.

3.7.11 RNA Fluorescence in situ hybridization (RNAscope) on intestinal organoids

For the RNAscope assay, organoids in matrigel were fixed in 4% 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.

3.7.12 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)[3,16] on the EPFL SCITAS HPC platform. After trimming and sequence tagging, reads were aligned to the human (hg38), mouse (GRCm38), or mixed reference genomes[155] (GSE63269), depending on the origin of the cellular input material, using STAR (version 2.7.0.e)[110]. 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 3.1H and Figure 3.1I). 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 3.1I) and for Drop-seq comparison experiments (Figure 3.1J) 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 3.1I) represent the following values: centerline, median; box limits, upper/lower quartiles; whiskers, 1.5x interquartile range; points, UMIs per cell.

3.7.13 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)[156]. 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. 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[138,157]. Pie-chart, bubble-plot and bargraph visualizations were carried out with *ggplot2*.

3.7.14 Slingshot analysis

The trajectories were constructed using the Slingshot wrapper implemented in the dyno package (https://github.com/dynverse/dyno)[157]. 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.

3.7.15 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 [140]. 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.

3.7.16 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 Palmer, 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[158]. 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[159] (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.

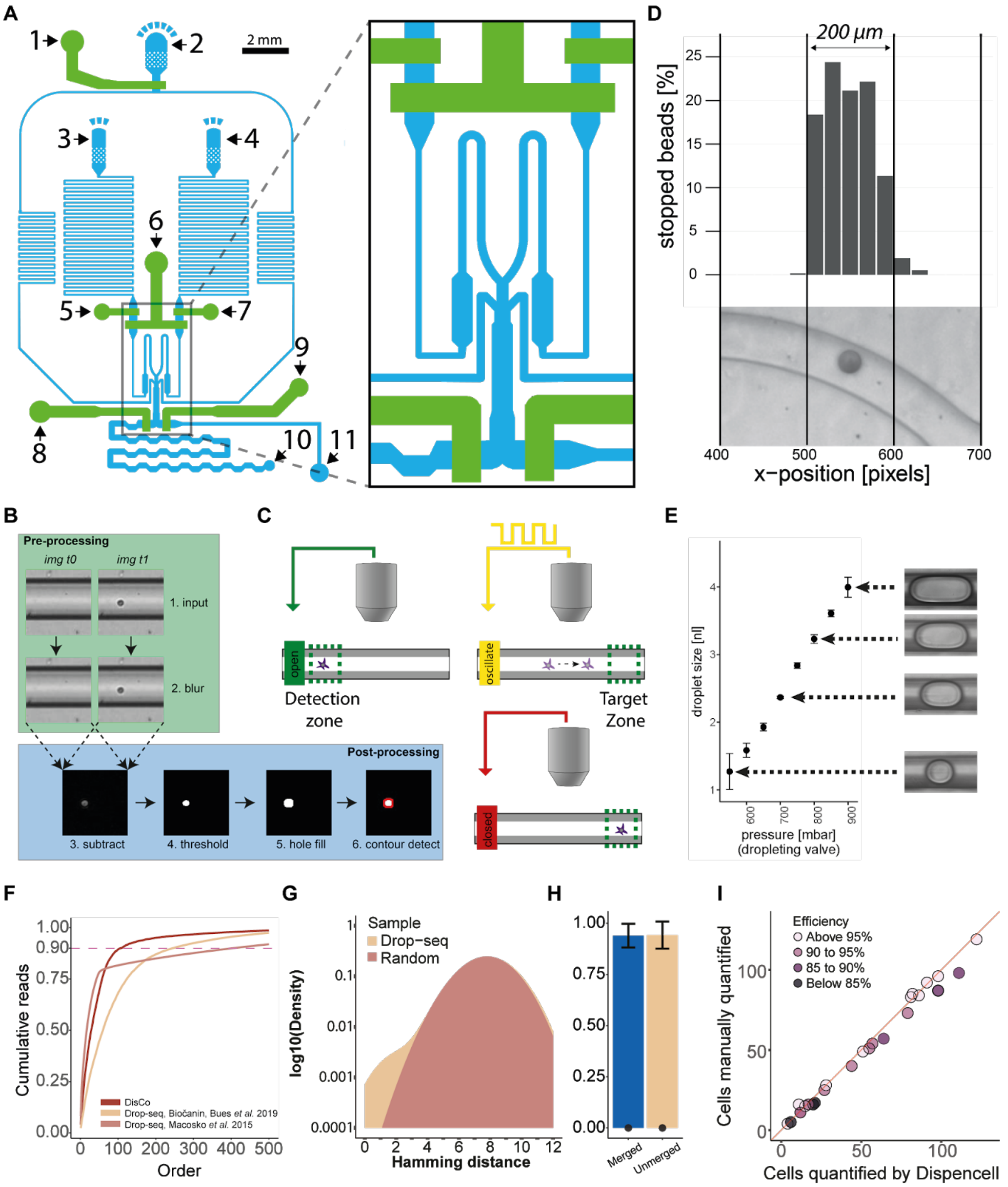
17. For intestinal organoid preparation for RNAscope, cold Cell Recovery Solution (Corning, 354253), Histogel (Thermo Scientific, HG-4000-012), Paraformaldehyde (4%, PFA, Electron Microscopy Sciences, 15714) were used.

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).

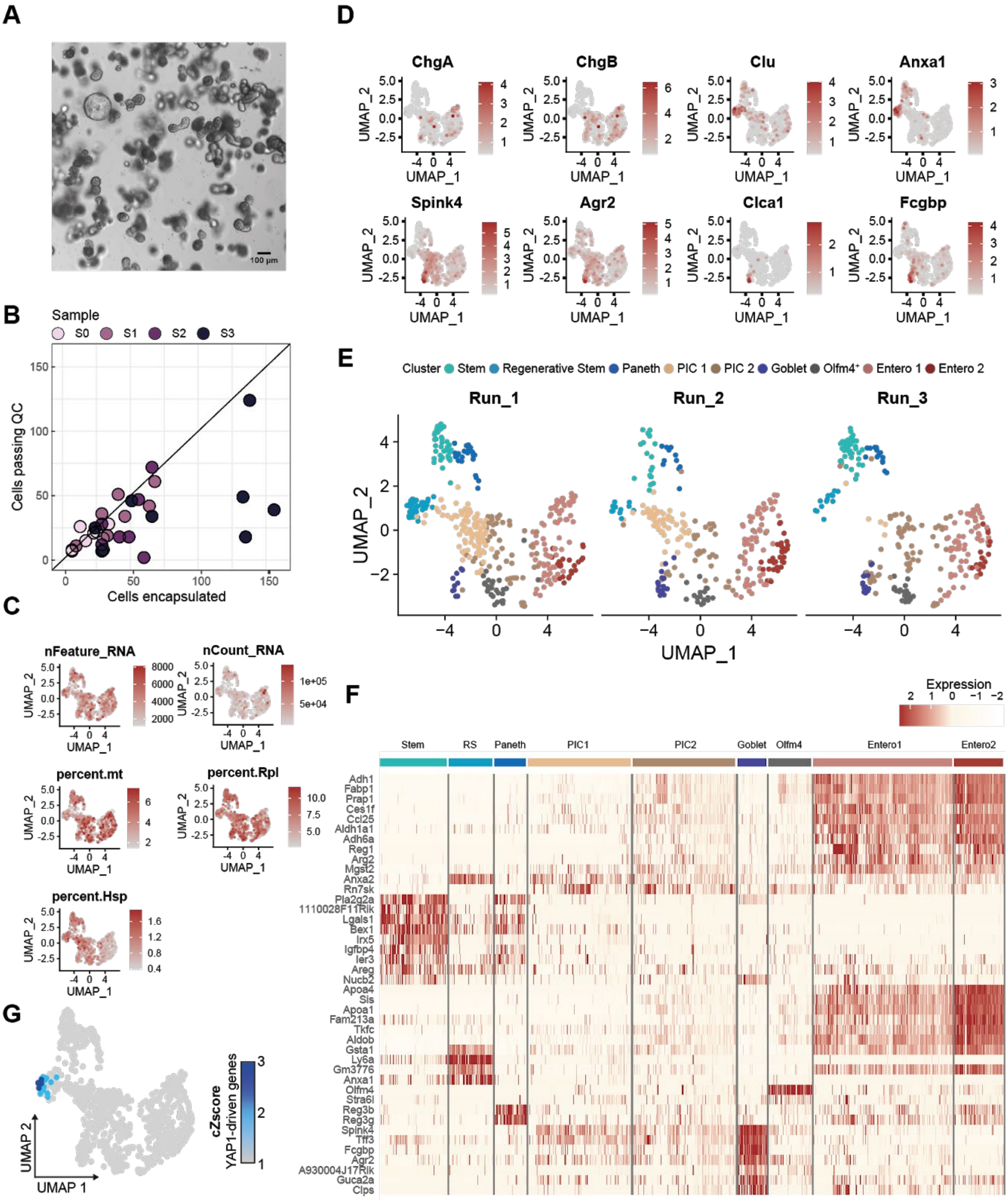
3.7.17 Data availability

The GEO accession number for scRNA-seq data reported in this paper is GSE148093.
The raw data for Figure 3.1H and Supplementary Figure 3.1H is stored under the access code GSM4454017. Corresponding digital gene expression (dge) summary data is available as the following file format: human.dge_summary.merged.txt, mouse.dge_summary.merged.txt. The raw data for Figure 3.1I and Supplementary Figure 3.1F is available under the access code GSM4454017. The corresponding dge summary files is human.dge_summary.merged.txt. The raw data for Figure 3.1J is stored under the access code GSM4454012 - GSM4454016, and the corresponding read count matrices *_ RCM_HD*.tar.gz, e.g. GSM4454016_RCM_HD1.tar.gz. The raw data for Figure 3.2, Supplementary Figure 3.2, Figure 3.3A, Figure 3.3E, Supplementary Figures 3.3C and Figure 3.3D is stored under access code GSM4453981- GSM4454011. Corresponding and read count matrices available under *_S*_*_RCMs.txt.gz, e.g. GSM4453981_S0_a_RCMs.txt.gz.

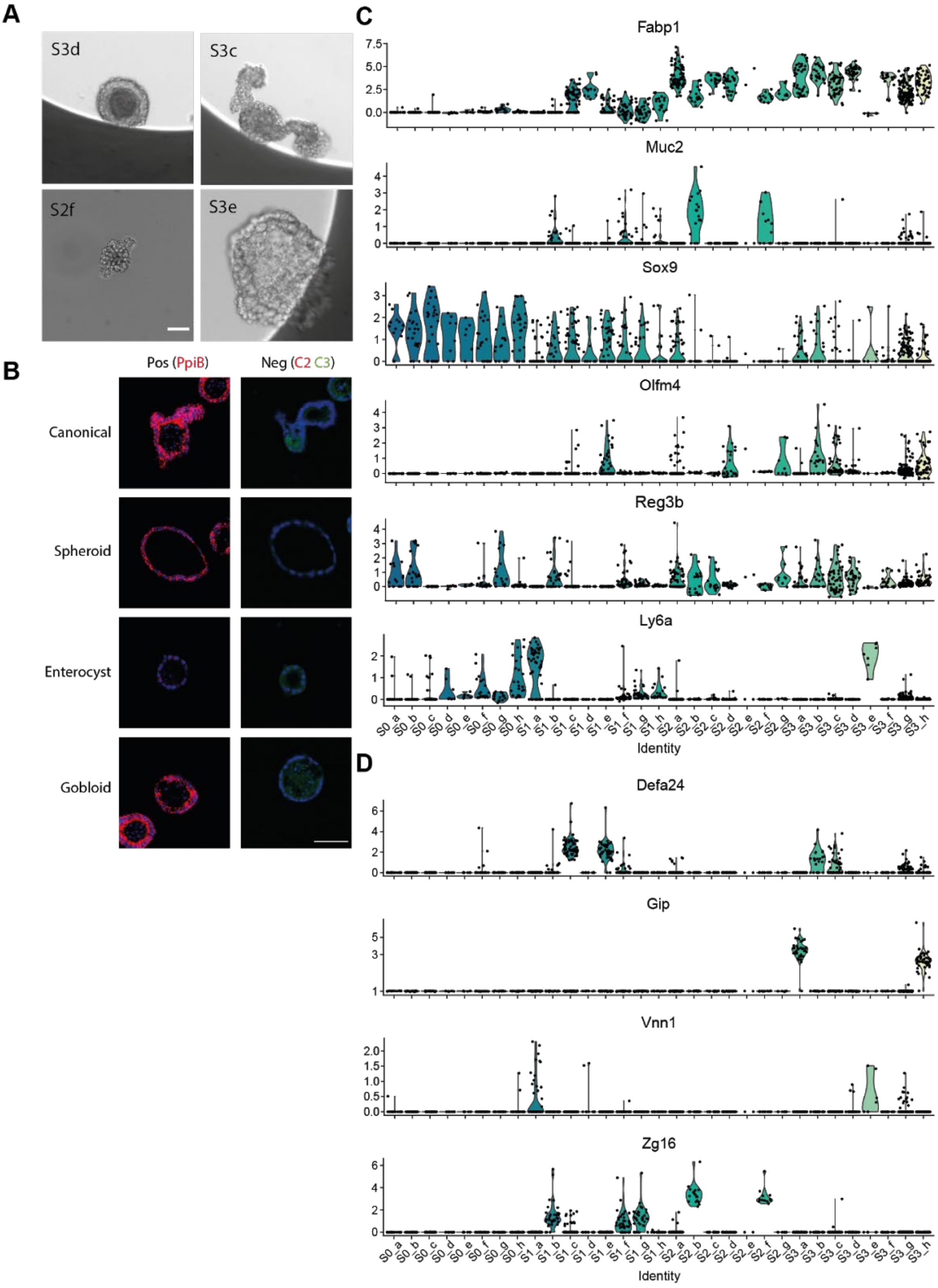
3.8. Supplementary Figures



Supplementary Figure 3.1 - Schematic of the DisCo workflow and performance metrics (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 despeckled 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 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[3,14]. **(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 standard deviation. **(I)** Correlation of the number of manually counted cells by fluorescence microscopy and the number of cells quantified by the DISPENCELL platform.



Supplementary Figure 3.2 - quality assessment of the scRNA-seq data generated using DisCo from intestinal organoids: **(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 number of detected UMIs per cell, number of detected genes per cell, percentage of mitochondrial reads, and 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)** Heatmap of top DE genes per annotated cluster. (**G**) YAP1 target gene activity on 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.



Supplementary Figure 3.3 - Individual intestinal organoid diversity: **(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 plot showing the expression of selected genes (*Defa24, Gip, Vnn1, Zg16*) identified via psupertime analysis per individual organoid.

3.9 Supplementary Table

(Left Subtable) Performance summary of established scRNA-seq platform technologies. Performance metrics were derived from literature. Noteworthy, as for lack of consensus experiments, metrics represent different values. (References and calculation of metrics are detailed in the Material and Methods section). (Right Subtable) Performance metrics calculated for the DisCo system are 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	1,000	50,000	10,000	10,000	200	1,600	400	< 50
Efficiency	45%*	25%*	2.3%*	-	-	55%**	43%**	30%*	75%*
$/cell (100 output cells)	$20	$2.1	$6	$10.6	$3.6	$29 (96 cells)	$5	$2.2	$1
$/cell (100 input cells)	$44.4	$8.4	$260.9	-	-	$62.2	$11.6	$7.5	$1.3
Additional remarks or limitations	Multiplexing possible, yet requires multiple washing procedures[57,160]. Substantial efficiency losses expected.			Fluorescent labeling necessary Expensive to scale up (automation)		Size-selective properties[39,145,]	High initial acquisition cost		

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

Chapter 4: Single-cell RNA-seq guided transcription factor gain of function screen in mesenchymal stem cells

Wanze Chen[123], Marjan Biočanin[123], Pernille Reiner[123], Wouter Saelens[123], Vincent Gardeux[12], Wangjie Liu[12], Julie Russeil[12], Radiana Ferero[12], Aly Makhlouf[12], Bart Deplancke[12]

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, Lausanne, Switzerland

3 Contributed equally to this work.

Shared First Author Contributions: WC, MB, PR and WS conceptualized the study, WC created backbone of barcoded vector, WC and MB cloned and prepared barcoded overexpression plasmids, WC and MB performed single-cell RNA-seq experiments, PR, WS, WC and MB performed the data analysis, MB performed confirmation experiments, MB, WC, PR and WS wrote the manuscript.

Chapter information: This chapter is a draft of the manuscript in preparation.

4.1 Abstract

Over recent years mesenchymal stem cells have gained traction as a promising model for basic and translation research due to their capacity to differentiate and transdifferentiate to cells of various lineages. The discovery of transcription factors that drive stem cell commitment towards different lineages was driven either by trial and error studies and more systematically by genetics screening. The latter are mostly in a “many for one” manner, testing many genes for one phenotypic or molecular readout. Here we have developed TF-seq by combining barcoded transcription factor-mediated cell programming and scRNA-seq method, a “many for many” strategy that allows to screen many genes for many molecular signatures that are defined by transcriptomic states. Applying TF-seq on mesenchymal stem cell-like C3H10T1/2 cells, we found multiple known factors that program the stem cells to adipose (such as *Cebpa*, *Pparg*), osteogenic (such as *Runx2*, *Dlx3*), and myogenic lineages (such as *Myog*, *Myod*). Northworthy, we discovered novel transcription factors, *Mycn* and *Rhox12* for adipogenesis, *Nkx3-1* and *E2f6* for osteogenesis, *Zfp687* and *Tbx21* for chondrogenesis. Further experimental validation confirms the function of *Mycn* and *Rhox12* genes in regulating adipogenesis, which together with previous results show the high efficiency and validity of the “many for many” strategy. Additionally, because TF-seq reads out the whole transcriptome, it allows for more causal studies on how the regulatory networks are affected by each TF overexpression. Finally, genome-wide changes driven by the transcription factors on the single-cell level render the possibility to study interactions between transcription factors and give deeper insights in cell-conversion efficiencies. We believe that data generated will serve as both experimental and data resource to the wider scientific community for multiplexed genetic assays.

4.2 Introduction

Multipotent mesenchymal stem cells (MSCs) are a promising resource for regenerative medicine, disease modeling and drug screening[73,161]. Under separate specific conditions, MSCs can differentiate into bone, fat, muscle, and cartilage cells[162–164]. Moreover, MSCs have a propensity to differentiate into cells outside the conventional mesenchymal lineages. Terminal differentiation of MSCs is achieved through coordinated activity of transcription factors (TFs) and epigenetic modifications that drive gene activity specific to distinct cell fates[164]. Current state of knowledge points out that these cells preserve a high degree of plasticity enabling trans differentiation between certain lineages. Overall, this makes MSCs ideal candidates for studying self-renewal, lineage commitment, differentiation and transdifferentiation both in *in vivo* and *in vitro* models[165–167]. One of the outstanding interests in both basic and translational stem cell research [168,169] is how to create cell type/state in a simple, controlled, and timely manner. This will allow further to dissect regulatory mechanisms governing the cell dynamics and transition through different states.

Perturbation of genes of interest and primary transcription factors has been shown as one of the key approaches in programming cell states. For example, it is shown that *Myog/Myod1* overexpression alone can reprogram fibroblasts to myoblasts through direct regulation of genes involved in myogenic gene expression programs[64,170]. In addition, expression of *Oct4*, *Sox2*, *Klf4* and *Myc* convert mouse embryonic fibroblasts to induced pluripotent stem cells[75]. The discovery of TFs that drive reprogramming has previously involved both prior knowledge of their role in development and trial and error experiments. A systematic screening method to assess the effects of TF overexpression would advance understanding on the contribution of individual TFs and enable rapid discovery of novel reprogramming factors and optimize already existing protocols.

Emergence of single-cell transcriptomics methods enabled for high-throughput genome-wide and high content screening in parallel, profiling thousands of individual transcriptomic responses to perturbation[11,12,171]. There are two most notable approaches to screening of perturbations in stem cells. First, multi-locus gene perturbation using clustered regularly interspaced short palindromic repeats (CRISPR/Cas9) allows both activation[172] and silencing of endogenous genes[77,173,174]. However, the efficiency of such methods remains limited as high concentration and multiple gRNA's are needed to achieve the silencing of one gene rendering the approach inefficient in some cases. Second, open-reading frame (ORF) overexpression has shown a potential to transdifferentiate cells between lineages and differentiate stem cells towards the desired state. ORF expression can yield a strong, stable expression of genes of interest[175]. Moreover, having full control on the ORF sequence enables expression of specific isoforms and mutants, and allows us to tightly regulate the gene's expression through inducible or context-specific promoters[80].

We developed TF-seq, a barcoded transcription factor ORF overexpression library, that we combine with scRNA-seq to perform gain-of-function screen in MSCs in a parallel fashion. Building on barcoding principles, we have developed a vector tagging method for reference lineages that helped us to delineate the MSC state lineage progression and differentiation towards adipo lineage. We generated a dataset where we probed for 435 TFs in ~52 000 cells and identified multiple candidate transcription factors in programming the cells into different cell types. Using building blocks from stem cell engineering and functional genomics, we aim to set up a framework for mechanistic understanding the role of TFs in designing cell states.

4.3. Results

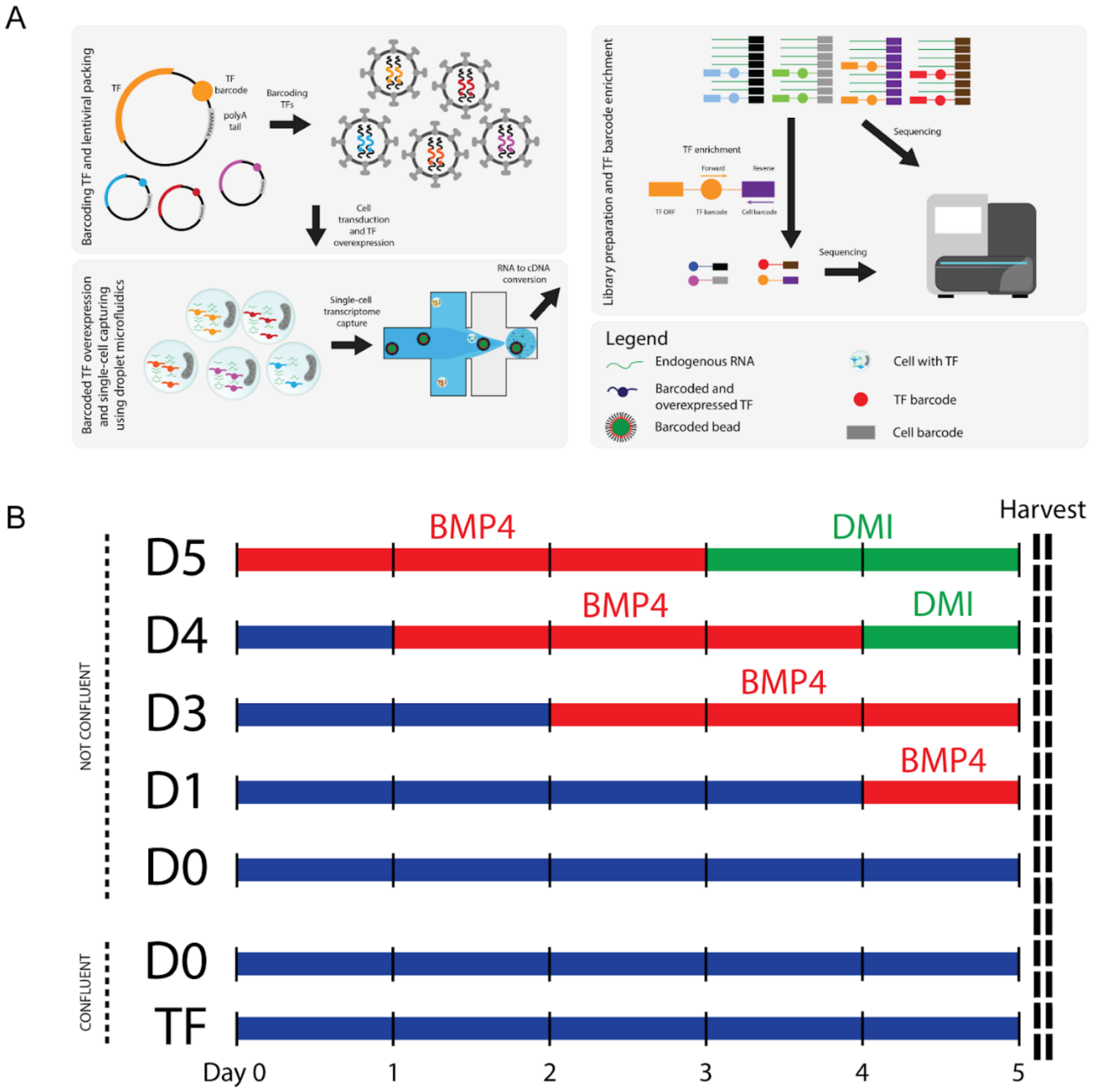

A
Barcoding TF and lentiviral packing
TF
TF barcode
polyA tail
Barcoding TFs
Cell transduction and TF overexpression
Barcoded TF overexpression and single-cell capturing using droplet microfluidics
Single-cell transcriptome capture
RNA to cDNA conversion
Library preparation and TF barcode enrichment
TF enrichment
Forward
Reverse
TF ORF
TF barcode
Cell barcode
Sequencing
Sequencing
Legend
Endogenous RNA
Barcoded and overexpressed TF
Barcoded bead
Cell with TF
TF barcode
Cell barcode
B
Harvest
D5
BMP4
DMI
D4
BMP4
DMI
D3
BMP4
D1
BMP4
D0
NOT CONFLUENT
D0
TF
CONFLUENT
Day 0
1
2
3
4
5

Figure 4.1 - Overview of theTF-seq principle: (A) Schematics of the TF-seq workflow. Barcoded ORF bearing plasmids are packed into lentiviruses. Following this, lentiviruses are transduced to C3H10T1/2 cells which are selected using puromycin. TF overexpression is selected is induced using doxycycline for 5 days, after which cells are dissociated and used for transcriptome analysis using 10X Chromium scRNA-seq platform. Once cDNA is prepared from single cells, part was used for single-cell library preparation and part was used for PCR enrichment of TF-ID and cellular barcode. Prepared libraries are then sequenced using Illumina platform. **(B)** Experimental design executed throughout the experiments. Next to overexpressed TF, reference dataset was created (D0 – D5) and D0 confluent cells that were cultured under the same conditions as overexpressed TFs

We generated a library containing 435 murine 3’ barcoded TF based on readily available TF Open Reading Frame (ORF) expression library available [94]. 11 bp random nucleotides (further called TF-ID) were added to the destination vector upstream of the 3’ UTR of the lentivirus vector to make the TF-ID detectable by 3' transcript capture method scRNA-seq. TF ORF was shuttled using Gateway cloning system to doxycycline inducible and puromycin selectable lentiviral vector. During cloning and barcoding procedure, expression vectors were thoroughly re-checked for uniqueness of TF-ID sequence and the right ORF. TF-ID of each vector was checked for similarity and those with similarity of less than 3bp were not included in the same experiment due to high likelihood of mixing. Lentiviruses were prepared individually avoiding recombination between different vectors during lentiviral packing as reported [176] (Figure 4.1A).

To generate TF programmed cells, we transduced C3H10T1/2 cells with the barcoded vector in individual manner. Lentivirus vectors containing *mCherry* were included as controls. Additionally, in order to generate a reference dataset of adipogenesis, we differentiated C3H10T1/2 cells in a time course manner (day1- 5) in parallel, with each time point labeled with a unique TF-ID. Instead of TF ORF, vectors used for reference datasets had included the *mCherry* red fluorescence gene (Figure 4.1B).

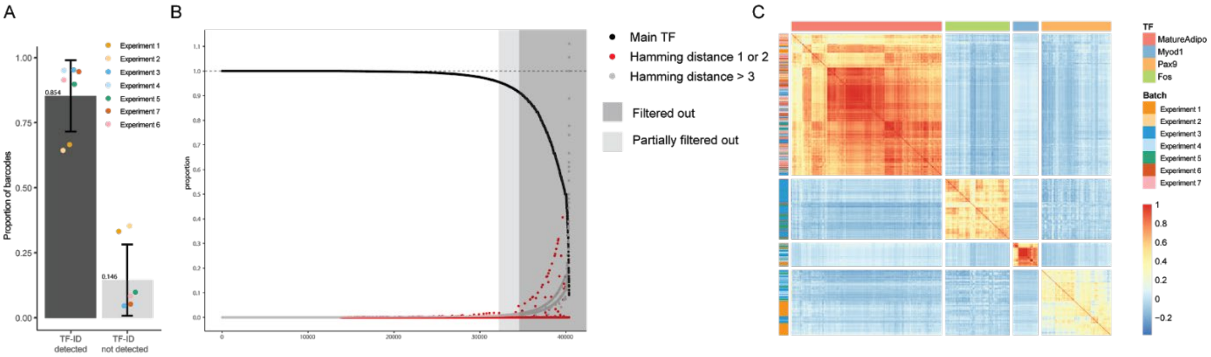


Figure 4.2 - TF-ID enrichment overview: (A) Barplot showing of TF-ID enrichment efficiency across different experiments, data points are colored according to the respective experiment **(B)** Cumulative plot showing proportion of reads assigned to unique TF-ID. Black dots represent the main detected TF-ID; red dots represent TF-IDs with hamming distance between 1 and 2 from main TF-ID; gray dots represent TF-IDs with hamming distance more than 3 from the main TF-ID **(C)** Heatmap showing similarity of gene-expression across several TFs repeatedly assayed across experiments.

With this design, we screened all the 435 TFs across seven experiments with 50 and 90 TFs per experiment and captured between 2000 and 8500 cells per one 10X scRNA-seq run (Supplementary Figure 4.1A) totaling ~52 000 cells with average coverage of 60 cells per TF. We observed lower gene/cell and UMI/cell detection in the first two experiments with 10X Chromium Next GEM Single Cell 3′ V2 chemistry compared to the V3 chemistry used in the other experiments (Supplementary Figure 4.1B). From the sequencing results of the conventional libraries, we could only recover the TF-ID of around 30-50% cells per run. We thus devised an additional enrichment strategy that was used to amplify the single-cell barcode and TF-ID generating additional libraries (termed as TF-ID enrichment library afterwards) sequenced along with the single-cell library (Figure 4.1A), which allowing on average 85% of all cells have been assigned with TF-ID (Figure 4.2A). Again, TF-ID assignments were somewhat lower in the first two experiments which is consistent with the previous observation of somewhat lower data quality in the first two experiments. We observed a strong correlation between the number of UMIs coming from vector expressions between scRNA-seq library and TF-ID enrichment library, supporting the accuracy of devised TF-ID enrichment approach. We expected to detect in only one TF-ID in one cell, as each cell only contains one TF. However, we found that around 40% of the cells contain more than one TF-ID (Figure 4.2B). There are several reasons that could lead to increased background: 1) the free-floating background RNA, 2) PCR or sequencing error, 3) cell doublets. We reasoned that the amount of false positive TF-IDs introduced by the first two types of artifacts are much less than the true-positive TF-ID in a cell, which can be corrected by filtering out the minor TFs (a cutoff at 5% of all the number of reads from TF-ID) and calculating the hamming distance of the TF-IDs, respectively (Figure 4.2B). The cells that could not be corrected by this approach are likely to be of the third type of artifact, and thus be removed for further analysis. To check for reproducibility of each experiment, we have repeatedly included several TFs (*mCherry* under adipogenic differentiation conditions, *Myod, Pax9, Fos*) across different experiments and observed strong clustering and similarity by TF or conditions over the experimental batches based on the transcriptomic state, hence reinforcing the robustness of overall experimental setup (Figure 4.2C).

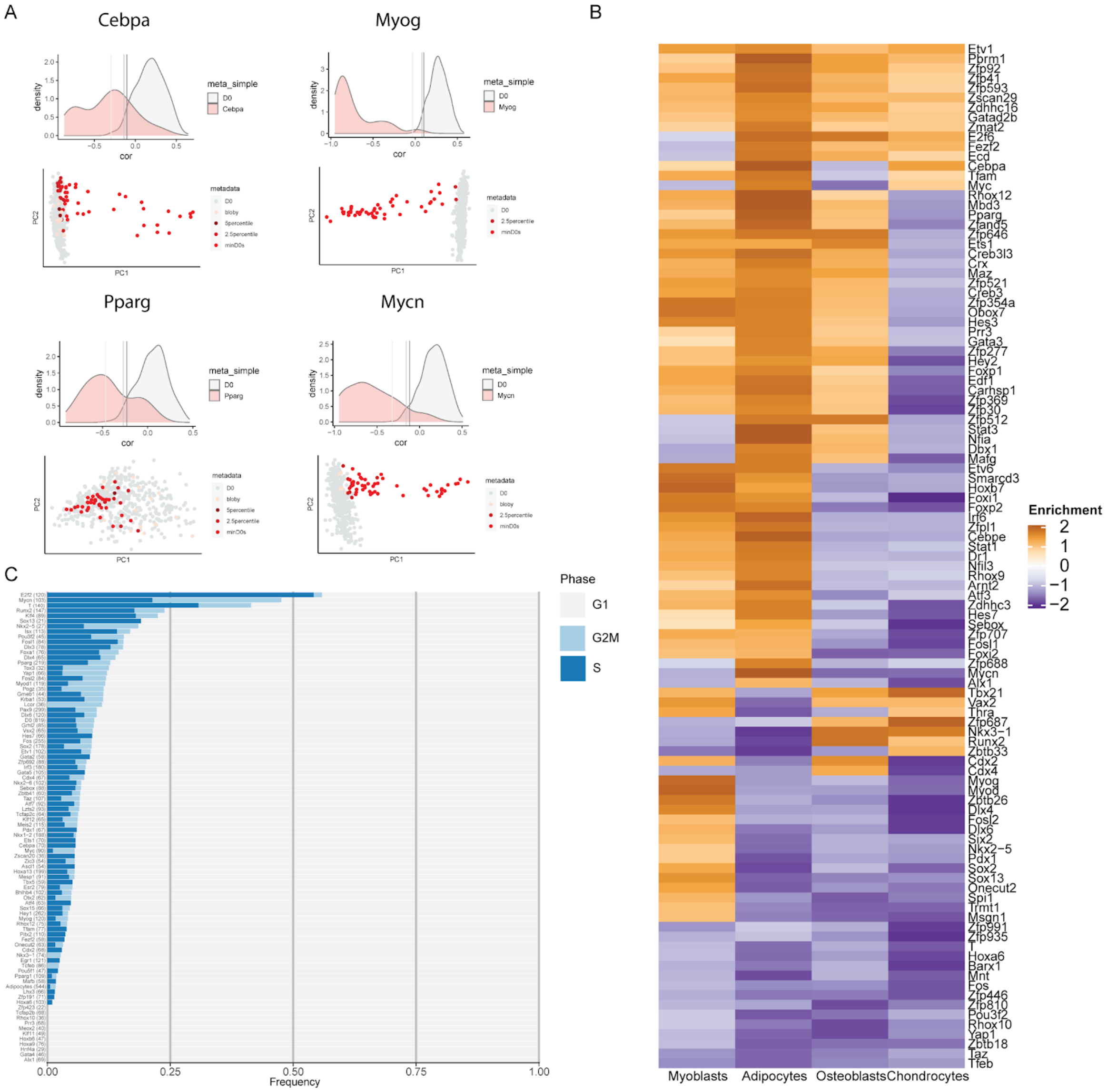


Figure 4.3 - Definition and characterization of functional cells in TF-seq framework (A) Functional cells of TFs driving myogenesis and adipogenesis. The top shows the distribution of correlation values in PCA space of all cells with the mean D0 cells. Functional cells are then defined as those that have a low correlation to all D0 cells. The bottom shows the location of the functional cells in the first two principal components. **(B)** Individual TFs that drive the main mesenchymal lineages. Shown are normalized enrichment values that were calculated using gene set enrichment analysis between marker genes of each lineage and the fold changes in gene expression induced by each TF. **(C)** Induction or repression of cell cycle of each TF. Shown are the % of cells that are assigned to the S phase (dark blue), G2M phase (light blue) or G1 phase (gray).

We next investigated which TFs are able to program the cells to adipocytes. We generate an adipose score using the adipose specific genes identified from the adipogenesis reference samples. By looking at the adipose score and comparing to the reference confirmed activity of several already known TFs such as *Cebpa*[177], *Pparg*[178]. Noteworthy we identified novel TF candidates potentially involved in adipogenesis such as *Rhox12* and *Mycn* (Figure 4.3B).

While we are at the moment in the process of adding more references such as osteogenesis and myogenesis reference datasets, currently available single cell atlas datasets nevertheless provide another valuable resource as references. By projecting our data to adipocyte, chondrocytes, osteoblasts and myocytes mouse references from PangloaDB [179],(Figure 4.2B) we further identified previously described TFs such as *Nkx3-1*, *Runx2* for chondrogenesis, *Runx2* for osteogenesis, *Myod* and *Myog* for myogenesis. To our surprise we found that some TFs previously described to be specific for certain lineages were enriched for more than one lineage. Such examples have already been described *Runx2* and *Pparg* , known as a master regulator for adipogenic differentiation[94,180,181]. Whether these different lineages share intermediate state during the early commitment needs to be further investigated.

In addition to characterizing the TF's function in programming different cell types, our TF-seq data is also a rich resource for other studies. This for example illustrated by the discovery of a set of TFs regulating cell cycle, including already known *E2f2*[182] , *Mycn*[183] and *T*[184]. It is possible to study the genetics interaction between different TFs, e.g. the strong correlation between some of the TFs within the same family showing the potential functional redundancy of these factors, such as *Fos, Hox and Myc, Cdx, Gata, Hox* and *Pou* (Supplementary Figure 4.2C). .

Some TFs-programmed cells show heterogeneity. One such example are *Pparg* programmed cells. We found that the level of vector expression (reflecting the TF expression) is a factor if not the main one causing the cellular heterogeneity. As a scRNA-based approach, TF-seq is able to exploit this heterogeneity to identify TFs which program cells in low efficiency and thus difficult to be identified in bulk based genetics screening. To identify TFs which show both strong and weak programming capacity (termed "functional TFs" afterwards), we compared the cells of each TF to those of control in single cell level. The outlier cells from a TF are defined as "functional cells" of that TF, which are further used, rather than all cells of that TF, to calculate the signature genes (DE genes) of this TF (see Materials and Methods). We then ranked the functionality of the TFs by those DE genes, including both strong and weak TFs .

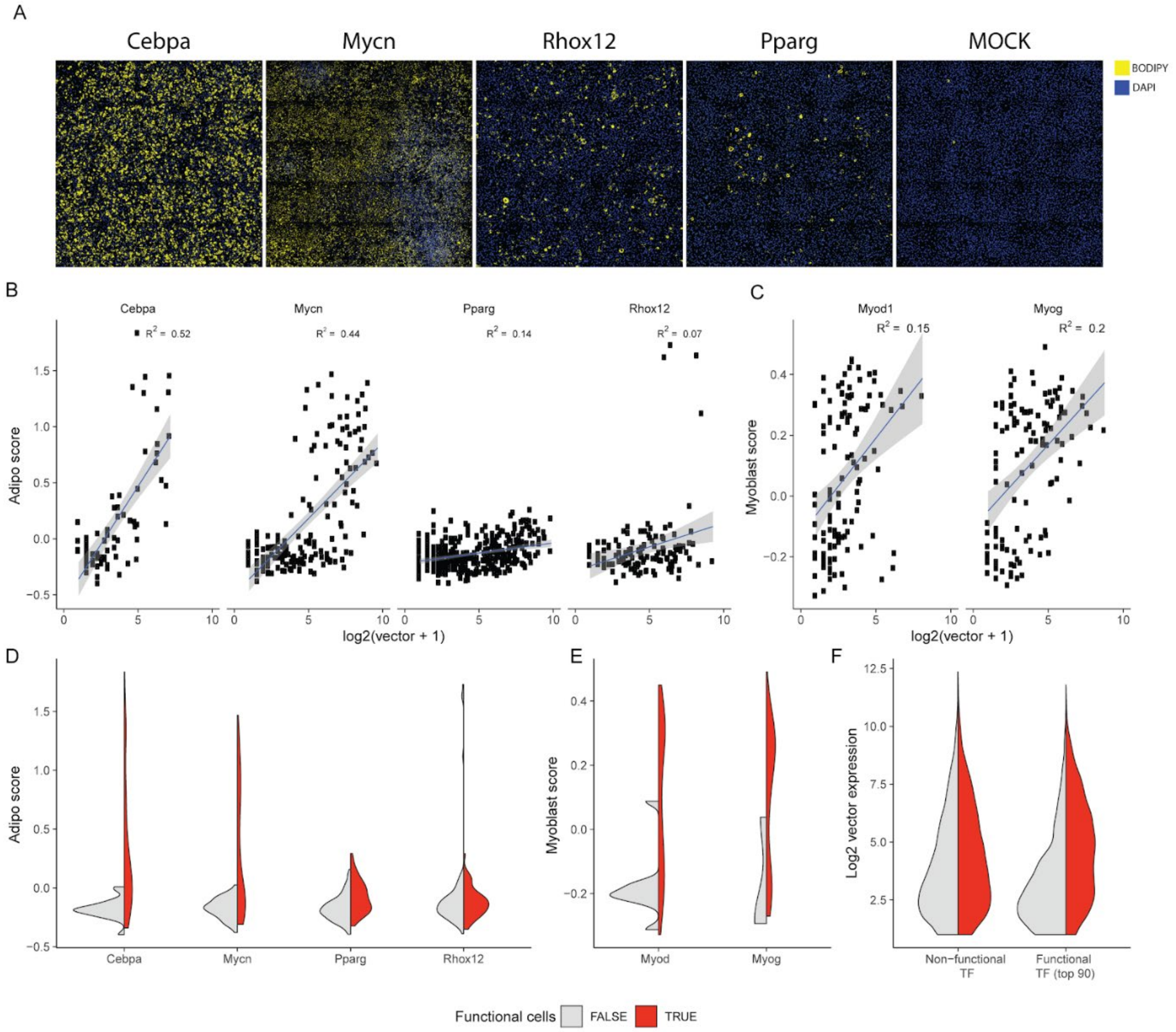


Figure 4.4 - Confirmation of TF activity and relationship to vector expression (A) Representative microscopy images after 7 day overexpression of *Cebpa, Mycn, Rhox12 Pparg,* and mock vector. After overexpression, cells were stained with DAPI and Bodpy and the whole wells were scanned using Operetta high-content analysis system. **(B-C)** Relationship between log vector expression and adipogenesis score (B) or myogenesis score **(D-E)** Relationship between functional cells and adipogenesis (D) and myogenesis score (E). **(F)** Relationship between functional cells and log vector expression. Top 90 functional TFs were selected based on having more than 200 differentially expressed genes.

We set to confirm whether we can phenotypically reproduce molecular profiles by prolonged overexpression (7 instead of 5 days). We confirmed that in cells overexpressed *Ceba. Mycn, Rhox12, Pparg* have varying degrees of lipid production, yet higher to mock group where mCherry was transduced. This initially confirmed our screening findings. Notable difference that we observed is higher cell density and different lipid droplet morphology compared to *Cebpa, Rhox12* and *Pparg*. To certain degree this is consistent with observations in initial screening where Mycn was among the top proliferating transcription factors while having pro-adipogenic activity (Figure 4.4A). As the vector level (reflecting the TF expression) can be quantified, TF-seq renders a kinetics analysis on the TF expression and the consequent molecular phenotypes. TFs that either induced an adipocyte or myocyte-like state for example showed a higher average expression of respective adipocyte and myoblast markers that was to some extent correlated with the log vector expression (Figure 4.4B-C). However, the amount of vector that was necessary to move the cell into a different state varied strongly between TFs. TFs such as *Myod* and *Myog* already induce a myocyte-like state at low vector expression which can be also observed by rapid cell morphology changes during overexpression, while TFs such as *Mycn* and *Rhox12* require a high vector expression for inducing adipocyte markers (Figure 4.4D-E).

4.4 Discussion

Here we developed barcoded lentiviral TF overexpression vectors used together with scRNA-seq to perform parallel gain-of-function screen in MSCs. This method allows not only for the discovery of strong drivers but also weak transcription factors that might have a significant contribution to defining both transitory and stable cell states. We discover and provide initial confirmations for novel transcription factors in adipogenesis – *Mycn* and *Rhox12*. Using the external references, we also identify several factors for other lineages. In this chapter we focused mostly on adipogenic lineage as a proof of principle, however, our method has the potential to focus on other lineages in a parallel manner with an aim to understand commitment and progression of MSCs towards defined cell types. We are currently in a phase building additional scaffolds for osteogenic and muscle lineages, which we believe will provide more insights than the external references, as they avoid the systematic variations of data coming from different studies. This will enable us to better annotate the changes within the already established framework of C3H10T1/2 differentiation.

One of the main drivers in our data variability was vector expression. We show that vector expression correlates with the level of transcriptomic response and strength of key lineage signatures. This goes in line with the initial results of the HumanTFome project where piggyBAC transposon vectors were chosen instead to achieve more efficient cellular programming[80]. However, such variability within vector expression gives us the ability to better estimate the dose of TF needed to successfully initiate the gene expression program required for successful lineage commitment and differentiation. Additionally, by leveraging on the vector expression needed to achieve the desired states in our cells, data presented here opens a new avenue to combinatorial overexpression of more than one transcription factor. Modeling such data will provide further insights in lineage commitment and further experimentally validate causal regulatory networks to achieve stable cell states.

4.5 Materials and Methods

4.5.1 TF barcoding and cloning

Barcoded inducible lentivirus plasmids containing transcription factor (TF) open reading frames (ORF) were created using the Gateway cloning system in two steps. In the first step, barcoded destination vectors were created. Barcoded lentiviral pTREP-vector was created by amplifying barcoded ccdB-containing barcoded fragment and vector-containing barcoded fragment. ccdB-containing fragment was amplified using pTREP-vector template (100 ng), Enrich_F3 primer (0.3 µM), pTREP-BC-RamR primer (0.3 µM), Kapa HiFi ready mix (1X, Roche, 07958935001) and dH2O following program: 1) 98°C for 3 min, 2) 98°C for 30 second, 3) 63°C for 30 second, 4) 72°C for 5 min, repeat 2-4 for 15 cycles, 5) 72 °C for 5 min. Vector containing fragment was amplified using pTREP vector template (100 ng), pTREP-vec-R primer (0.3 µM), pTREP-BC-RamF primer (0.3 µM), Kapa HiFi ready mix (1X, Roche, 07958935001) and dH_2O following program 1) 98°C for 3 min, 2) 98°C for 30 second, 3) 63 °C for 30 second, 4) 72°C for 5 min, repeat 2-4 for 15 cycles, 5) 72 °C for 5 min.. Following this, both PCR fragments were run on 1% Agarose gel and purified using Zymo gel purification kit (Zymo, D4007). Once purified, pTREP vector fragments were assembled using Gibson assembly mix (NEB, E2611S) using ccdB-containing barcoded fragment (100 ng), pTREP vector-containing barcoded fragment, Gibson assembly mix (1X, NEB, E2611S) at 50°C for 60 minutes. Reaction was then purified using DNA Clean and Concentrator purification kit (Zymo, D4014) and the final product was transformed using One Shot ccdB Survival 2 T1R resistant competent cells (Thermo Fisher, A10460) following manufacturer's instructions. Successful colonies were then inoculated to LB containing ampicillin and chloramphenicol for miniprep (Zymo, D4015) and validation. In the second step, transcription factor open reading frames were transferred from entry pDONR clones generated (paper reference) to pTREP vector using LR Clonase II enzyme mix (Thermo Fisher, 11791100) following manufacturer's instructions. Stbl3 one shot competent cells (Invitrogen, C737303) were then transformed using LR Clonase II enzyme mix (Thermo Fisher, 11791100) and colonies were grown on ampicillin plates (100 ug/mL) overnight. Colonies were then picked and transferred to LB with ampicillin for miniprep. Following miniprep, plasmids were checked for TF-ID presence and correct TF ORF using Sanger sequencing (Microsynth) using EGFP-C-Rev standard primer to confirm TF-ID, TET-CMV-for standard primer to confirm TF. Prior to each scRNA-seq experiment, TF-IDs pooled together were checked for similarity and only TF-IDs with hamming distance dissimilarity greater than 2 bp were retained within each pool.

4.5.2 Cell culture

Both HEK 293T (ATCC Cat. No. SD-3515) and C3H10 T1/2 cells (ATCC Cat. No. CCL-226) were cultured in DMEM (Gibco, 41966029) supplemented with FBS (10%, Gibco, 10270106) and 1X penicillin-streptomycin (1%, Life Technologies, 15140-122) in a 5% CO2 humidified atmosphere at 37 °C. Prior to use, cells were washed with PBS (Thermo Fisher, 14190169), dissociated with Trypsin-EDTA (0.25%, 25200056, Life Technologies), washed with cell wash buffer and counted with Trypan blue (Life Technologies, T10282) live-dead stain using a Countess cell counter (Invitrogen).

4.5.3 Transfection and lentiviral packing

Reverse transfection was performed using Lipofectamine 2000 (LP2000, Thermo Fisher, 11668027) following manufacturers instruction. First, a mix of Opti-MEM (Thermo Fisher, 31985070) and LP2000 reagent was prepared and thoroughly mixed. Next, mix of Opti-MEM, TF bearing lentiviral vector, (1.5 µg) and 3rd generation lentivirus packaging plasmids (pRSV-Rev:pMDLg/pRRE:pCMV-VSV-G=1:1:1, 1.5 µg). Two mixes were mixed together and incubated for 30 minutes at the room temperature. During incubation, HEK 293T (ATCC Cat. No. SD-3515) cells were washed with PBS (Thermo Fisher, 14190169), then dissociated using Trypsin-EDTA (0.25%, 25200056, Life Technologies). Cells were resuspended in DMEM (Gibco, 41966029) FBS (10%, Gibco, 10270106) and 1X penicillin-streptomycin (1%, Life Technologies, 15140-122) and seeded in individual wells at ~95% confluency. After the incubation, the mix was added to the medium containing freshly seeded cells. 12 hours post-transfection, fresh medium was added to attached cells. 48 hours post transfection, medium was collected, and dead cells were removed, centrifuging at 300 x g. As a control, pBOB-GFP (kindly provided by Dr. Jiahuai Han in Xiamen University) plasmid with constitutive GFP expression was transfected as a control to evaluate success of transfection step.

4.5.4 Transduction and overexpression

Murine C3H10 T1/2 cells were seeded 12 hours prior to transduction at density 10-20%. Then, transfection lentivirus containing supernatant mixed in 1:1 ratio with fresh medium and polybrene (10 µg/mL, Sigma, TR-1003-G) was added. Cells in a plate were then centrifuged at 1300 x g for 30 minutes (37°C,). Cells were incubated with lentivirus for 24h and then in fresh medium. After 48h, cells were selected using puromycin (2 µg/mL, Thermo Fisher, A1113803) for 72 hours. After puromycin was removed, fresh medium was added to allow puromycin resistant cells to recover for 24 to 48h. Then, medium containing doxycycline was added (2 µg/mL, Sigma, D9891-1G) to induce TF overexpression. Fresh medium with doxycycline was replenished every 48h.

4.5.5 In vitro adipogenic differentiation

For building adipose reference differentiation dataset, selected cells with unique TF-ID and *mCherry* ORF, were cultured in the DMEM (Gibco, 41966029) supplemented with *FBS* (10%, Gibco, 10270106) and 1X penicillin-streptomycin (1%, Life Technologies, 15140-122) , and BMP4 (100 mg/mL, RnD, 314-BP-010) for 3 days. Then the induction medium was added for 2 days, which is the complete culture medium supplemented with the MDI cocktail (1 µM dexamethasone, D4902-25MG, 0.5mM 3-isobutyl-1-methylxanthine, I5879-5g, and 167 nM insulin, I9278-5ml, all from Sigma-Aldrich, Saint Louis, MO). Teh cells were then kept in complete medium with insulin (167 nM insulin, Sigma-Aldrich I9278-5ml) until collection.

4.5.6 10X scRNA-seq experiment

scRNA-seq experiments were performed using the Chromium Single Cell Gene Expression Solution (10X Genomics, v2 reaction kit for experiment 1-2 and v3 reaction kit for 3-7 experiment), following the manufacturer's protocol. After 5 days of doxycycline induced overexpression, cells were trypsinized and recovered from plates. Following this, cells were washed and resuspended in PBS (Thermo Fisher, 14190169). Cell suspension (concentrated to 500 – 800 cells/µL) was added to Chromium Single Cell 3' chip targeting 6000 - 8000 cells per experiment. cDNA and single-cell RNA-seq libraries were prepared using Chromium Single Cell 3' v2 and v3 reagent kits. Libraries were sequenced on the NextSeq 500 (Illumina) instrument using the dual-index configuration following manufacturer's recommendations to obtain mean depth 50 000 reads/cell.

4.5.7 TF-ID enrichment

10 ng of cDNA generated from each Chromium Single Cell 3' experiment was used to enrich TF-ID to recover overexpressed transcription factor identity. Enrichment was performed in two steps. In the first step, target TF-ID was amplified in PCR reaction using ~6 ng cDNA, BC_vec_target_10X_F1 vector specific forward primer (0.3 µM), Trueseq universal adaptor (0.3 µM), Kapa HiFi ready mix (1X, Roche, 07958935001) and dH_2O following 98°C for 30s, 10 cycles of 98°C for 10s, 63°C for 20s, 72°C for 30s, followed by 5 min elongation at 72°C. Enriched cDNA was then purified using Ampure beads (2.5:1 ratio, Labgene, CNGS-0050) and purified DNA was further amplified with Truseq_D7_adaptor (0.3 µM),

Truseq_Universal_Adapter (0.3 μM), Kapa HiFi ready mix (1X) following the program 98ºC for 30s, 4 cycles of 98ºC for 10s, 63ºC for 20s, 72ºC for 30s, followed by 5 min elongation at 72ºC to make the libraries compatible with 10X libraries. Following this, TF-ID enriched libraries were then purified twice using Ampure beads (0.6:1 ratio) and sequenced together with 10X libraries.

4.5.8 Data preprocessing

For each experiment, only the cell-barcodes detected both in the 10X and TF-ID library were used. The TF-IDs for each cell barcode were ranked based on the percentage of reads assigned in the TF-ID library. The TF-ID of a cell was defined as the one with the highest fraction of reads. The cells were then ordered based on the percentage of reads assigned to their main TF-ID, and knee point detection was performed to find the inflexion points of the curve using the function kneepointDetection() of the R package SamSPECTRAL[185]. The cells with a percentage lower than the knee point were filtered out.

Prior to integrating the datasets from different experiments, each dataset was analysed individually to filter low quality cells. Outlier cells were filtered out using the median absolute deviation from the median total UMI numbers and gene numbers, implemented in the function isOutlier() of package SCRAN[186], using a cutoff of 4 or 6 depending on the dataset. Cells with more than 10% of mitochondrial genes, 40% of rRNA genes and less than 75% of protein coding genes were filtered out.

4.5.9 Assigning cells to TF-IDs

We created the vector sequence as a *.fasta* file, leaving 'N' where the TF-ID barcodes were located. Then, we aligned the R2 *.fastq* files from 10X TF-ID enriched libraries to the vector sequence using STAR v2.7.3a[110]. Using custom scripts, we extracted the barcodes from the aligned reads at the 'Ns' locations, and matched them to our barcode libraries, allowing 1nt mismatch. Using the R1 *.fastq* file, we were then able to match the TF-IDs to the corresponding cell barcodes. This created a TF-IDs x cell barcodes count matrix, that we further deduplicated using UMIs from the R1 *.fastq* file.

4.5.10 Selection of functional cells based on PCA distance to D0 cells

Functional cells were selected for each TF based on their distance to the mean of D0 and D0 confluent groups in the PCA space. More precisely, the cells were first assigned to a cell cycle phase (G1, G2/M, S) using the function CellCycleScoring() of Seurat[187]. The threshold of the score to assign a cell to a certain phase was modified to 1 (instead of 0 by default), meaning that a cell was assigned to the phase between G2/M and S with the highest score only if it was greater than 0.1, the cell was otherwise assigned to G1. The rest of the selection of functional cells analysis was performed per phase. To limit batch effect, the selection was performed per datasets, however to have a sufficient number of D0 cells, D0 from datasets experimentally prepared and sequenced together were merged. A PCA was computed for a TF's cells and the corresponding D0 cells. The PCA was computed on normalized and scaled data and on the 2000 highly variable features as implemented in Seurat[187]. The mean expression of D0 cells (per gene) was calculated on the first two principal compontents (PCs) and the other significant PCs, with a p value smaller than 0.05 as calculated by Jackstraw. Pearson correlation between each cell and the mean D0 was calculated using cor() function. TF's cells with a correlation smaller than the correlation corresponding to the 2.5 percentile of the correlation distribution of D0's cells with their own mean were considered as functional.

4.5.11 Data integration

The TFs selected for the integration were defined as enriched in a specific cluster of the analysis per batch. More specifically clustering was performed following the Seurat pipeline, and a resolution was chosen so that it represented the heterogeneity of the data. For each TF and each cluster with specific resolution significant enrichment was tested using Fisher's exact test. By doing this, we filtered 90 TFs. All the cells of filtered TFs, cells defined as mature adipocytes, and D0 cells of each batch were selected for the further integration. A new Seurat object of the subset of selected cells per batch was created to log normalize the data and find the top 2000 highly variable features. The features used for the integration were the union of the top 2000 highly variable features of each of these Seurat objects, the genes corresponding to the 90 functional TFs, and the top 10 differentially expressed markers of the specific clusters where the TFs of interest was also included in the further analysis. The integration was finally performed using mnnCorrect() function from the R package batchelor[188]. The integrated normalized gene expression was then scaled in order to compute the PCA. UMAP and clusters were computed on the first 150 PCs using Seurat[156]. TF enrichment in a specific cluster of the integration was calculated as described previously. The Heatmap of figure 4.2C and Supplementary figure 4.2C show the Pearson correlation of the selected cells in the first 150

PCs. The cells displayed are the subset of cells of the TFs of interest that were in the cluster representative of the TFs of interest as defined by Fisher's exact test.

4.5.12 Determining candidate transcription factors that drive adipogenesis and other lineages

We used three different strategies to determine candidate TFs that drive a particular lineage. To define individual TFs that drive adipogenesis, we calculated a score for each TF's functional cells based on the core set of adipogenesis genes as defined earlier (using Seurat's ModuleScore function). This allowed us to rank TFs by which they promote or repress adipogenesis.

To calculate differential expression during adipogenesis or TF overexpression, we used a generalized linear model with batch effect and % of non-zeros as covariates[189,190]. For adipogenesis, we compared the day zero confluent cells with the mature cell population. For TF overexpression, we compared both day zero confluent and non-confluent cells with the functional cells for a particular TF.

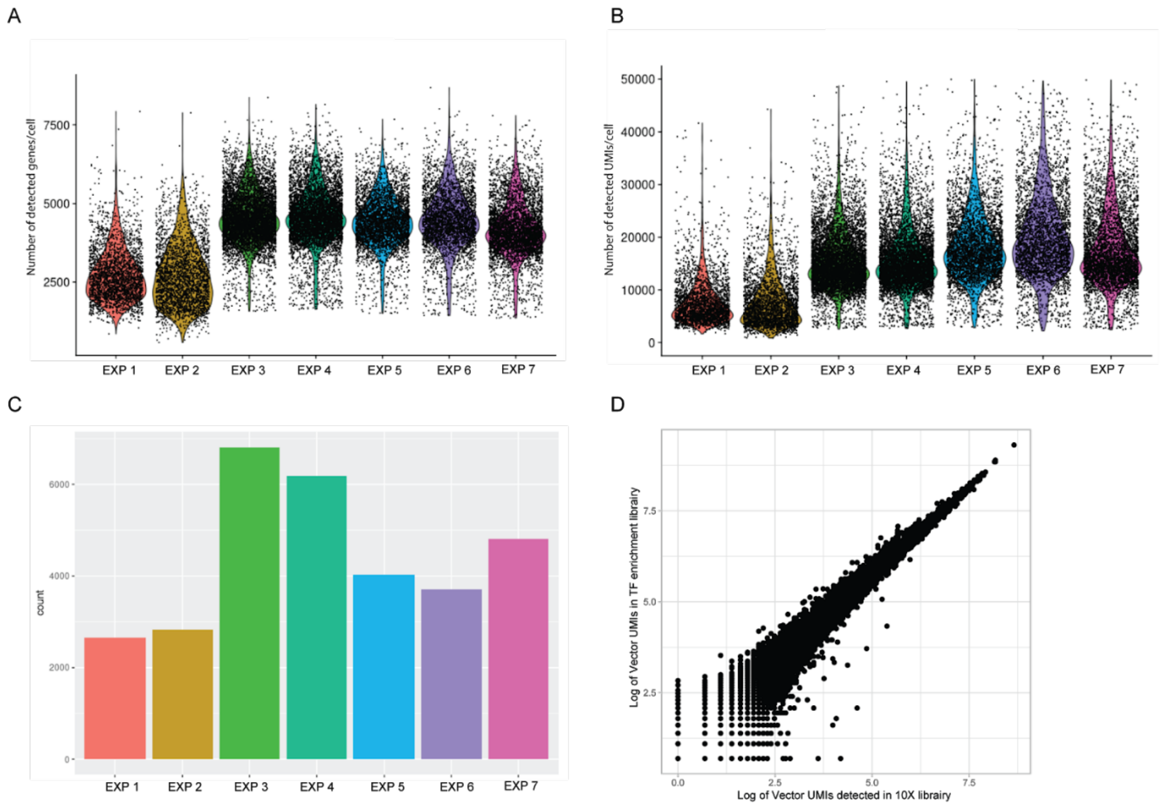
To define combinations of TFs that may drive adipogenesis, we then constructed a regularized linear model that tries to predict the log fold-changes of the adipogenesis differential expression using a sparse combination of the log fold-changes of TFs. Specifically, we calculated for each combination of two to four TFs the most optimal set of regression coefficients using least-squares, and used these to calculate a mean-squared error value between the combination of TFs.

To define individual TFs that may drive lineages beyond adipogenesis, we used PanglaoDB[179] to extract markers for the four main mesenchymal lineages: adipogenesis, myogenesis, osteogenesis and chondrogenesis. We then used gene-set enrichment analysis, as implemented in the R fgsea package[191], to calculate p-values and normalized enrichment scores between the gene set and fold-changes of a particular TF. Candidate TFs were selected if they were significantly enriched or depleted in at least one gene set (false-discovery rate < 0.05).

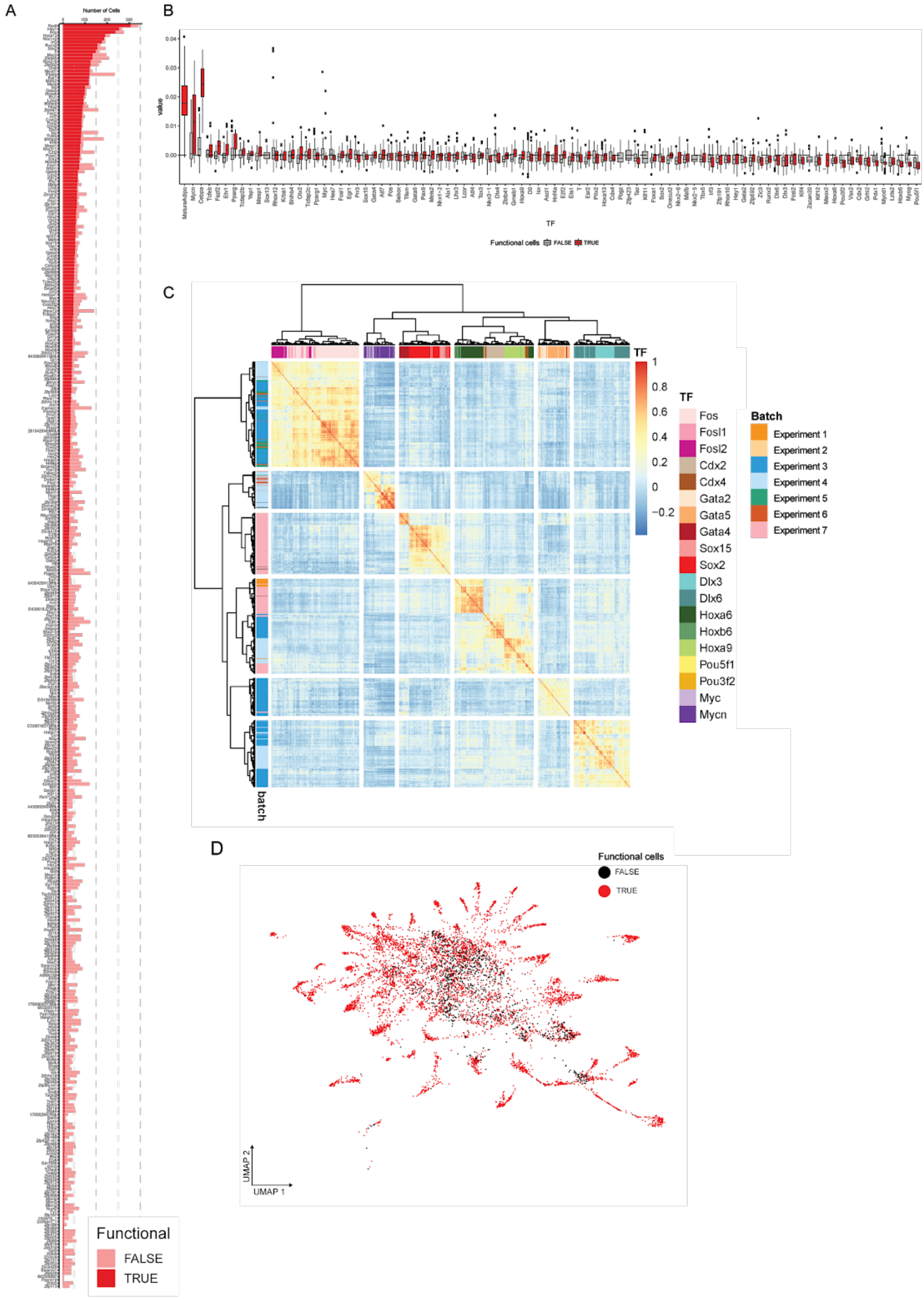
4.5.13 Confirmation of candidate TF driven adipose differentiation

To confirm candidates for adipogenesis, TFs of interest in three technical replicates were overexpressed for 7 days in C3H10 T1/2 cells in the flat bottom imaging plate (Corning Falcon, 353219). Following this, cells were washed with PBS (Thermo, 14190169) and fixed with 4% PFA (Electron Microscopy Sciences, 15714) in plate for 15 minutes at room temperature in dark. Cells were washed twice with PBS and then permeabilized with PBS and triton (0.3%, Applichem, A4975.0100) followed by two more steps of washing with PBS. To stain the lipid droplets, cells were incubated in DAPI (1:5000, Sigma-Aldrich, D9542-1MG) and Bodipy (1:500, Life Technologies, D3922) for 30 minutes in the dark at room temperature. Cells were then washed twice with PBS and imaging was performed in PBS. Imaging was performed using Operetta high-content imaging system (Perkin Elmer). Stack of 10 images per whole well (96 well plate) was collected for each replicate for both blue and green channel using 20x/0.8 objective. To estimate differences in adipose differentiation driven by overexpressed TFs, image pre-processing and quantification algorithm developed by EPFL BIOP imaging facility (available through EPFL BIOP ImageJ repository) was used.

4.6 Supplementary figures



Supplementary Figure 4.1 - Quality parameters of individual scRNA-seq experiments (A) Violin plot showing number of detected genes/cell across different experiments **(B)** Violin plot showing the distribution of detected UMIs/cell across different experiments **(C)** Number of cells retained per experiment after initial quality filtering and matching to respective TF-ID **(D)** Correlation of overexpression vector UMIs/cell detected from 10X Chromium scRNA-seq library and TF-ID enrichment.



Supplementary Figure - 4.2 cell pre-filtering and functional cells with significant TF overexpression (A) Number of functional vs non-functional cells for each TF across batches. **(B)**

Adipogenesis score of functional vs non-functional cells for each TF and the mature adipocyte population. **(C)** Correlation in PCA space between cells with TFs that were replicated across batches. **(D)** Dimensionality reduction of functional TFs, with functional cells highlighted.

Chapter 5: Discussion and outlook

5.1 Method development in single-cell transcriptomics domain

As the adoption of microfluidic methods for various biological applications broadens, hardware setup requirements will have to evolve towards setups that are able to efficiently process cells at high-throughput, yet, remain broadly adoptable with ease of use allowing for a standardization among use cases.

With the development of e-chip and cp-chip for Drop-seq, we were able to modularize and extend the overall setup that can be utilized for various Drop-seq approaches, allowing for broader adoption in the community. With the novel cp-chip, we achieved highly efficient sample recovery at the pivotal methodological step, particularly when cell and bead numbers are scarce. Importantly, by introducing an additional microfluidic process in the Drop-seq workflow, the novel bead processing workflow simplified cDNA generation and without hampering the data quality as the original setup.

With DisCo we developed a promising prototype platform, that, unlike Drop-seq, inDrops, and 10X[32,43,102,132] approaches, lowers the bar in terms of minimal sample requirements while remaining very efficient in sample utilization. Although, establishing DisCo in other labs around the world will require effort, the benefits of this initial investment may be outweighed by the profits similar as for other widely used microfluidic technologies such as Drop-seq if sole focus of the lab requires routine handling samples that are not easy to process using the other approaches such as FACS based scRNA-seq or 10X scRNA-seq solutions. Considering that non-commercial alternative technologies are similarly complex in operation and equipment cost, we are convinced that the DisCo approach will prove valuable for the processing of small and precious biological samples at the single-cell level. Importantly, DisCo thereby fills an important processing gap left by currently available methods[59] for mid-small samples of 100 to 500 cells with recoveries above 70%. In order to leverage this unique feature, set of high recovery and small sample numbers, we studied early-stage development heterogeneity in individual intestinal organoids. By dissecting single organoids at single-cell resolution, we unveiled four different subtypes of organoids at various stages of development. Compared to high-content screening methods that rely on specific marker genes to study development, we were able to uncover so far undescribed organoid type enriched with goblet cells, here thermed "gobloids" and confirmed existence of two more aberrant forms described elsewhere[138,192] by using the single-cell transcriptomics approach agnostic to specific markers[122,192].

As a standalone technology, DisCo is currently facing four challenges in the near future. Firstly, the throughput of DisCo is limited at maximally 300 cells/hour. As DisCo was initially developed to fill an open niche in the scRNA-seq toolbox[43,101,103,153,193], we aimed to provide the means to efficiently process rare and low input samples using conventional dual-layer microfluidics. Importantly, such PDMS-based systems have excessive latency times, making placement of multiple particles per encapsulation process a time-expensive endeavor for bigger samples. Hence, we aim to reduce those latency times, by either removing the necessity for complete opening and closing of the valves or developing a system that would not rely on PDMS-based valves in the future, both of which have the potential to increase throughput. The second major major limitation to throughput resides in the coordination of two particles before co-encapsulation. At the point of technology creation, Drop-seq beads[32] were adopted due to their ease of use and coordination on a microfluidic device and presumably acceptable production standards. However, removing the requirement to coordinate bead movement will decrease the time significantly. One approach would be to establish closed-packed ordering[194] of beads utilized in 10X and inDrops[102]. Alternatively, particle co-encapsulation could be replaced with barcoded liquid biochemistry, avoiding the utilization of RNA capture beads. This leads us to the third challenge - data quality. Since the release of the original Drop-seq protocol, the field has moved foward and away from solid Drop-seq beads as RNA capture strategy[52]. One of the main reasons for this is the large variation in bead quality production and low RNA to cDNA conversion on the bead's surface resulting in high-dropout rates. The most widely used system to date relies on the replacement of solid RNA capture beads with dissolvable hydrogel beads[102,195], prominently featured by 10X genomics, or alternatively liquid biochemistry based on the golden standards such as Smart-seq[23,193,196] or SCRB-seq[197]. Data quality would significantly improve by either utilizing dissolvable hydrogel beads or implementing an alternative barcoding system in the framework of liquid biochemistries already existing for plate based scRNA-seq approaches. The fourth major challenge but also opportunity for further developments rests in the enhanced utilization of machine-vision previous and during the co-encapsulation process. By embedding the DisCo microfluidics platform in a substantially more sophisticated microscopy setup, the already integrated machine-vision could not only be used for profiling cells, but also could pave the way towards multi-parametric information linking cell phenotype and transcriptional state, hence connect two amongst most prominent approaches to high-throughput cellular profiling. Such development would allow for even further sub-sorting cells of interests (i.e., based on fluorescence signal) from heterogeneous samples to further characterize the population of interest on a full transcriptome level. One such example would be to study the difference between fully-differentiated cells and the cell populations that fail to reprogram during the TF induction within a target stem cell population.

5.2 Single-cell perturbation assays

Furthermore, we established TF-seq, a barcoded transcription factor open-reading frame overexpression library combined with scRNA-seq to perform gain-of-function screening in murine MSC-like *in vitro* model. Unique vector barcoding of ORFs makes this method compatible with other scRNA-seq approaches and other genome-wide approaches like. scATAC-seq[198]. Framework for cell-perturbation presented here can be easily applied to any other murine models (i.e. gastruloids, ESCs). We envision that this library will be further expanded, ultimately covering all known mouse protein TFs. This is to the current knowledge, the largest barcoded murine TF screen performed on the single-cell level and second largest after human TFome project[80]. Additionally, utilizing TF-seq approach we found multiple known factors involved in programming stem cells towards, adipose, osteogenic and myogenic lineages and discovered novel factors that might be potentially involved in mentioned processes.

Whilst establishing TFseq we identified several opportunities to extend current progress. Firstly, tight control over the virus titer would allow for a more fine-grained grasp on TFs expression and thereby enable to understand how the level of expression yields novel cell state. As mentioned above, potential technological advances in scRNA-seq toolbox would allow to fully profile cellular phenotype together with transcriptomic program. This also opens an avenue to more accurate TF activity quantification through direct fluorescent labeling (specifically using panel of HA-tags to measure TF activity) of overexpressed TF prior to RNA sampling.

Second, by fully harnessing the ability to express TF upon induction, one can also achieve more precise temporal control enabling to perform assays in conjugation with environmental stimulation of cells. There are numerous examples where tight temporal control of TF activity is detrimental to cell fate decisions[192,199–201].

Third, integration of available TF binding data with single-cell data presented in this thesis will further aid in understanding whether how much TF expression is needed to achieve the desired perturbation of gene-expression profile thus desired cell state. Furthermore, using the wealth of scRNA-seq data from single TF overexpression, this approach paves a way towards combinatorial gene perturbations. By looking at the individual target genes expression in the context of vector expression at the single-cell level, interaction partners[66], binding sites across the genome[69,202], binding affinities, we will be able to generate further mechanistic insights that orchestrate gene-regulatory networks and drive cellular heterogeneity.

5.3 Future perspectives

Both the technological and biological developments presented in this thesis should aid broader adoption of single-cell technologies and assays. Yet, several open common challenges remain to be addressed in the future. For most tissue biopsies, but also the used organoids, a denominating challenge is to efficient is a high yield dissociation thus developing addressing this problem with further open single-cell transcriptomics to analysis of hard-to-process samples.

Even if high yields of single cells from tissue precious samples will be attainable, a core argument of single cell suspensions remains, being that due to cell dissociation, tissue structure and interactions are lost. Several bioinformatics approaches tried to overcome this by predicting ligand interactions[203,204]. Another more permanent approach is the emergence of spatial transcriptomics over the past years. Approaches like Visium or Slide-seq [205]allowed for profiling thousands of genes on a tissue slice by preserving original tissue structure and cellular interactions. However, current resolution of these methods is still in the ranges between 50 and 100 µm, meaning that these methods are still beyond the single-cell resolution whilist more targeted approaches like smFISH still remain attractive yet limited to in number of markers utilized[206]. An additional argument against using current single-cell methods is that they provide a snapshot of cellular state at the moment of sampling. Independently developed, quantitative imaging can overcome these limitations by dynamically following transcript production and decay[207]. However, these analyses are usually limited to several genes of interest. I envision that future developments will further build on these principles – continuous assaying of genome-wide gene expression without killing or significantly disrupting the normal cell activity. In the end, with the further developments of multi-omics assays and assays to sample genome, proteome and transcriptome of the same cell, we will be able to integrate all systems fully approaches at the single-cell level.

Finally, as broad adoption by the biology community increases, I believe that all techniques discussed here will further evolve to be fully integrated into standard biology research that yields a better understanding of multicellular interactions in the context of immune responses, development and adipogenesis. At the moment concerted efforts evolve around the completion of major atlasing projects[11,55,113] (i.e. Fly Cell Atlas, Mouse Cell Atlas and Human Cell Atlas) and the inherent data integration and interpretation challenges encompassing different transcriptome- and genome-wide approaches. Upon completion of fully characterizing cell types through their molecular state and lineages data from such projects will become the starting point for hypotheses development, yet it will not be sufficient alone. If to fully understanding the cellular functions in tissue and organismal-specific context, single-cell transcriptomics will only progress in conjugation with established life sciences fields like (oncology, immunology, developmental biology), further adding single-cell omics approaches to the already available toolbox within the respective fields. Currently, common fishing expeditions and explorative research will become a stricter exclusivity of new technologies and methods emerging in the field, including approaches like single cell proteomics[208]. In contrast, hypothesis-driven research will have to harness the full power of the methods established at a faster pace to properly integrate into the existing biological frameworks.

References

1. Crick, F. Central dogma of molecular biology. *Nature* (1970) doi:10.1038/227561a0.
2. Hinman, V. & Cary, G. The evolution of gene regulation. *Elife* (2017) doi:10.7554/elife.27291.
3. Arendt, D. *et al.* The origin and evolution of cell types TL - 17. *Nat. Rev. Genet.* **17** , 744–757 (2016).
4. Tolar, J., Le Blanc, K., Keating, A. & Blazar, B. R. Concise review: Hitting the right spot with mesenchymal stromal cells. *Stem Cells* (2010) doi:10.1002/stem.459.
5. Bluguermann, C. *et al.* Novel aspects of parenchymal-mesenchymal interactions: From cell types to molecules and beyond. *Cell Biochem. Funct.* (2013) doi:10.1002/cbf.2950.
6. Man, K., Kutyavin, V. I. & Chawla, A. Tissue Immunometabolism: Development, Physiology, and Pathobiology TL - 25. *Cell Metab.* **25** , 11–26 (2017).
7. Sotelo, C. Viewing the brain through the master hand of Ramon y Cajal. *Nat. Rev. Neurosci.* (2003) doi:10.1038/nrn1010.
8. Howard Gest. Homage to Robert Hooke (1635–1703): New Insights from the Recently Discovered Hooke Folio. *Perspect. Biol. Med.* (2009) doi:10.1353/pbm.0.0096.
9. Kawai, J. *et al.* Functional annotation of a full-length mouse cDNA collection. *Nature* (2001) doi:10.1038/35055500.
10. Dunham, I. *et al.* An integrated encyclopedia of DNA elements in the human genome. *Nature* (2012) doi:10.1038/nature11247.
11. Regev, A. *et al.* The human cell atlas. *Elife* (2017) doi:10.7554/eLife.27041.
12. Svensson, V., Vento-Tormo, R. & Teichmann, S. A. Exponential scaling of single-cell RNA-seq in the past decade. *Nature Protocols* vol. 13 599–604 (2018).
13. Coffin, J. M. & Fan, H. The Discovery of Reverse Transcriptase. *Annual Review of Virology* (2016) doi:10.1146/annurev-virology-110615-035556.
14. Bustin, S. A. Absolute quantification of mrna using real-time reverse transcription polymerase chain reaction assays. *Journal of Molecular Endocrinology* (2000) doi:10.1677/jme.0.0250169.
15. Velculescu, V. E., Zhang, L., Vogelstein, B. & Kinzler, K. W. Serial analysis of gene expression. *Science (80-.).* (1995) doi:10.1126/science.270.5235.484.
16. Goodwin, S., McPherson, J. D. & McCombie, W. R. Coming of age: Ten years of next-generation sequencing technologies. *Nature Reviews Genetics* (2016) doi:10.1038/nrg.2016.49.
17. Tang, F. *et al.* mRNA-Seq whole-transcriptome analysis of a single cell. *Nat. Methods* (2009) doi:10.1038/nmeth.1315.
18. Kellogg, R. a, Gómez-Sjöberg, R., Leyrat, A. a & Tay, S. High-throughput microfluidic single-cell analysis pipeline for studies of signaling dynamics. *Nat. Protoc.* **9**, 1713–26 (2014).

19. Islam, S. *et al.* Quantitative single-cell RNA-seq with unique molecular identifiers. *Nat. Methods* **11**, 163–6 (2014).

20. Hashimshony, T., Wagner, F., Sher, N. & Yanai, I. CEL-Seq: Single-Cell RNA-Seq by Multiplexed Linear Amplification. *Cell Rep.* **2**, 666–673 (2012).

21. Jaitin, D. A. *et al.* Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. *Science (80-.).* (2014) doi:10.1126/science.1247651.

22. Picelli, S. *et al.* Smart-seq2 for sensitive full-length transcriptome profiling in single cells TL - 10. *Nat. Methods* **10** , (2013).

23. Hagemann-Jensen, M. *et al.* Single-cell RNA counting at allele and isoform resolution using Smart-seq3. *Nat. Biotechnol.* (2020) doi:10.1038/s41587-020-0497-0.

24. White, A. K. *et al.* High-throughput microfluidic single-cell RT-qPCR. TL - 108. *Proc. Natl. Acad. Sci. U. S. A.* **108** , 13999–14004 (2011).

25. Folch, A. *Introduction to BioMEMS*. *Introduction to BioMEMS* (2016). doi:10.1201/b12263.

26. Guo, M. T., Rotem, A., Heyman, A. & Weitz, D. A. Lab on a Chip Droplet microfluidics for high-throughput biological assays {. **i**, 2146–2155 (2012).

27. Haeberle, S. & Zengerle, R. Microfluidic platforms for lab-on-a-chip applications. *Lab Chip* **7**, 1094–1110 (2007).

28. Sesen, M. & Whyte, G. Image-Based Single Cell Sorting Automation in Droplet Microfluidics. *Sci. Rep.* (2020) doi:10.1038/s41598-020-65483-2.

29. Debs, B. El, Utharala, R., Balyasnikova, I. V., Griffiths, A. D. & Merten, C. A. Functional single-cell hybridoma screening using droplet-based microfluidics. *Proc. Natl. Acad. Sci. U. S. A.* (2012) doi:10.1073/pnas.1204514109.

30. Liu, H. *et al.* A Droplet Microfluidic System to Fabricate Hybrid Capsules Enabling Stem Cell Organoid Engineering. *Adv. Sci.* (2020) doi:10.1002/advs.201903739.

31. Lagus, T. P. & Edd, J. F. High-throughput co-encapsulation of self-ordered cell trains: cell pair interactions in microdroplets. *RSC Adv.* **3**, 20512 (2013).

32. Macosko, E. Z. *et al.* Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* **161**, 1202–1214 (2015).

33. Klein, A. M. *et al.* Droplet Barcoding for Single Cell Transcriptomics Applied to Embryonic Stem Cells. 1–24.

34. Teh, S.-Y., Lin, R., Hung, L.-H. & Lee, A. P. Droplet microfluidics. *Lab Chip* **8**, 198–220 (2008).

35. Chong, Z. Z. *et al.* Active droplet generation in microfluidics. *Lab Chip* (2015) doi:10.1039/C5LC01012H.

36. Klein, A. M. & Macosko, E. InDrops and Drop-seq technologies for single-cell sequencing TL - 17. *Lab Chip* **17** , 2540–2541 (2017).

37. Klein, A. M. *et al.* Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* (2015) doi:10.1016/j.cell.2015.04.044.

38. Ziegenhain, C. *et al.* of Single-Cell RNA Sequencing Methods. **65**,.

39. DeLaughter, D. M. The Use of the Fluidigm C1 for RNA Expression Analyses of

Single Cells. *Curr. Protoc. Mol. Biol.* (2018) doi:10.1002/cpmb.55.

40. Xin, Y. *et al.* Use of the Fluidigm C1 platform for RNA sequencing of single mouse pancreatic islet cells. *Proc. Natl. Acad. Sci. U. S. A.* (2016) doi:10.1073/pnas.1602306113.

41. Kolodziejczyk, A. A., Kim, J., Svensson, V., Marioni, J. C. & Teichmann, S. A. The Technology and Biology of Single-Cell RNA Sequencing TL - 58. *Mol. Cell* **58 VN-r**, 610–620 (2015).

42. Ziegenhain, C. *et al.* Comparative analysis of single-cell RNA-sequencing methods. **49**,.

43. Zheng, G. X. Y. *et al.* Massively parallel digital transcriptional profiling of single cells TL - 8. *Nat. Commun.* **8** , (2017).

44. Systems, T. S. R. *et al.* Comparative Analysis of Droplet-Based Ultra-High- Article Comparative Analysis of Droplet-Based. *Mol. Cell* 1–13 (2019) doi:10.1016/j.molcel.2018.10.020.

45. Packer, J. S. *et al.* A lineage-resolved molecular atlas of C. Elegans embryogenesis at single-cell resolution. *Science (80-.).* (2019) doi:10.1126/science.aax1971.

46. 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.

47. Couturier, C. P. *et al.* Single-cell RNA-seq reveals that glioblastoma recapitulates a normal neurodevelopmental hierarchy. *Nat. Commun.* (2020) doi:10.1038/s41467-020-17186-5.

48. Weinreb, C., Rodriguez-Fraticelli, A., Camargo, F. D. & Klein, A. M. Lineage tracing on transcriptional landscapes links state to fate during differentiation. *Science (80-.).* **367**, (2020).

49. Morris, S. A. The evolving concept of cell identity in the single cell era. (2019) doi:10.1242/dev.169748.

50. Wang, Y. J. *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.

51. Zhang, X. *et al.* Comparative Analysis of Droplet-Based Ultra-High-Throughput Single-Cell RNA-Seq Systems. *Mol. Cell* (2019) doi:10.1016/j.molcel.2018.10.020.

52. Svensson, V. *et al.* Power analysis of single-cell RNA-sequencing experiments TL - 14. *Nat. Methods* **14**, 381–387 (2017).

53. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species TL - 36. *Nat. Biotechnol.* **36** , 411 (2018).

54. Schaum, N. *et al.* Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. *Nature* (2018) doi:10.1038/s41586-018-0590-4.

55. Davie, K. *et al.* A Single-Cell Transcriptome Atlas of the Aging Drosophila Brain TL - 174. *Cell* **174**, 982–805306368 (2018).

56. Isakova, A., Fehlmann, T., Keller, A. & Quake, S. R. A mouse tissue atlas of small noncoding RNA. *Proc. Natl. Acad. Sci. U. S. A.* (2020) doi:10.1073/pnas.2002277117.

57. McGinnis, C. S. *et al.* MULTI-seq: sample multiplexing for single-cell RNA sequencing

using lipid-tagged indices. *Nat. Methods* (2019) doi:10.1038/s41592-019-0433-8.

58. Stoeckius, M. *et al.* Simultaneous epitope and transcriptome measurement in single cells TL - 14. *Nat. Methods* **14**, (2017).

59. Hwang, B., Lee, J. & Bang, D. Single-cell RNA sequencing technologies and bioinformatics pipelines TL - 50. *Exp. Mol. Med.* **50** , (2018).

60. Wu, H. *et al.* Dual functions of Tet1 in transcriptional regulation in mouse embryonic stem cells. *Nature* (2011) doi:10.1038/nature09934.

61. Wilkinson, A. C., Nakauchi, H. & Göttgens, B. Mammalian Transcription Factor Networks: Recent Advances in Interrogating Biological Complexity. *Cell Systems* vol. 5 319–331 (2017).

62. Peñalosa-Ruiz, G., Bright, A. R., Mulder, K. W. & Veenstra, G. J. C. The interplay of chromatin and transcription factors during cell fate transitions in development and reprogramming. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms* vol. 1862 194407 (2019).

63. Vaquerizas, J. M., Kummerfeld, S. K., Teichmann, S. A. & Luscombe, N. M. A census of human transcription factors: Function, expression and evolution. *Nature Reviews Genetics* (2009) doi:10.1038/nrg2538.

64. Tapscott, S. J. The circuitry of a master switch: Myod and the regulation of skeletal muscle gene transcription. *Development* (2005) doi:10.1242/dev.01874.

65. Zhou, Q. *et al.* A mouse tissue transcription factor atlas. *Nat. Commun.* (2017) doi:10.1038/ncomms15089.

66. Aibar, S. *et al.* SCENIC: Single-cell regulatory network inference and clustering. *Nat. Methods* (2017) doi:10.1038/nmeth.4463.

67. Guo, C. & Morris, S. A. ScienceDirect Engineering cell identity : establishing new gene regulatory and chromatin landscapes. *Curr. Opin. Genet. Dev.* **46**, 50–57 (2017).

68. Hens, K. *et al.* Automated protein-DNA interaction screening of Drosophila regulatory elements TL - 8. *Nat. Methods* **8** , 1065–1070 (2011).

69. Isakova, A. *et al.* SMiLE-seq identifies binding motifs of single and dimeric transcription factors. *Nat. Methods* (2017) doi:10.1038/nmeth.4143.

70. Rockel, S., Geertz, M. & Maerkl, S. J. MITOMI: A microfluidic platform for in vitro characterization of transcription factor-DNA interaction. *Methods Mol. Biol.* (2012) doi:10.1007/978-1-61779-292-2_6.

71. Friman, E. T. *et al.* Dynamic regulation of chromatin accessibility by pluripotency transcription factors across the cell cycle. *Elife* (2019) doi:10.7554/eLife.50087.

72. Scheres, B. Stem-cell niches: Nursery rhymes across kingdoms. *Nature Reviews Molecular Cell Biology* (2007) doi:10.1038/nrm2164.

73. Pittenger, M. F. *et al.* Mesenchymal stem cell perspective: cell biology to clinical progress. *npj Regenerative Medicine* vol. 4 1–15 (2019).

74. Biddy, B. A. *et al.* Single-cell mapping of lineage and identity in direct reprogramming TL - 564. *Nature* **564** , 219–224 (2018).

75. Takahashi, K. & Yamanaka, S. Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors TL - 126. *Cell* **126** , 663–676 (2006).

76. Dixit, A. *et al.* Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens TL - 167. *Cell* **167** , 1853 (2016).

77. Adamson, B. *et al.* A Multiplexed Single-Cell CRISPR Screening Platform Enables Systematic Dissection of the Unfolded Protein Response TL - 167. *Cell* **167** , 1867 (2016).

78. Replogle, J. M. *et al.* Combinatorial single-cell CRISPR screens by direct guide RNA capture and targeted sequencing. doi:10.1038/s41587-020-0470-y.

79. Nguyen, Q. H. *et al.* Single-cell RNA-seq of human induced pluripotent stem cells reveals cellular heterogeneity and cell state transitions between subpopulations. *Genome Res.* (2018) doi:10.1101/gr.223925.117.

80. Ng, A. H. M. *et al.* A comprehensive library of human transcription factors for cell fate engineering. *Nat. Biotechnol.* (2020) doi:10.1038/s41587-020-0742-6.

81. Duan, J. *et al.* Rational Reprogramming of Cellular States by Combinatorial Perturbation. *Cell Rep.* **27**, 3486-3499.e6 (2019).

82. Pradhan, R. N. *et al.* Dissecting the brown adipogenic regulatory network using integrative genomics TL - 7. *Sci. Rep.* **7**, 42130 (2017).

83. Schwalie, P. C. *et al.* A stromal cell population that inhibits adipogenesis in mammalian fat depots. *Nature* (2018) doi:10.1038/s41586-018-0226-8.

84. Sidossis, L. & Kajimura, S. Brown and beige fat in humans: Thermogenic adipocytes that control energy and glucose homeostasis. *Journal of Clinical Investigation* (2015) doi:10.1172/JCI78362.

85. Bartelt, A. & Heeren, J. Adipose tissue browning and metabolic health. *Nature Reviews Endocrinology* (2014) doi:10.1038/nrendo.2013.204.

86. Majka, S. M. *et al.* De novo generation of white adipocytes from the myeloid lineage via mesenchymal intermediates is age, adipose depot, and gender specific. *Proc. Natl. Acad. Sci. U. S. A.* (2010) doi:10.1073/pnas.1003512107.

87. Macotela, Y. *et al.* Intrinsic differences in adipocyte precursor cells from different white fat depots. *Diabetes* (2012) doi:10.2337/db11-1753.

88. Morin, X., Daneman, R., Zavortink, M. & Chia, W. A protein trap strategy to detect GFP-tagged proteins expressed from their endogenous loci in Drosophila TL - 98. *Proc. Natl. Acad. Sci.* **98**, 15050–15055 (2001).

89. Wu, J. *et al.* Beige adipocytes are a distinct type of thermogenic fat cell in mouse and human. *Cell* (2012) doi:10.1016/j.cell.2012.05.016.

90. Pradhan, R. N. *et al.* Dissecting the brown adipogenic regulatory network using integrative genomics. *Sci. Rep.* (2017) doi:10.1038/srep42130.

91. Long, J. Z. *et al.* A smooth muscle-like origin for beige adipocytes. *Cell Metab.* (2014) doi:10.1016/j.cmet.2014.03.025.

92. Cristancho, A. G. & Lazar, M. A. Forming functional fat: A growing understanding of adipocyte differentiation. *Nature Reviews Molecular Cell Biology* (2011) doi:10.1038/nrm3198.

93. Rosen, E. D. & MacDougald, O. A. Adipocyte differentiation from the inside out. *Nature Reviews Molecular Cell Biology* (2006) doi:10.1038/nrm2066.

94. Gubelmann, C. *et al.* Identification of the transcription factor ZEB1 as a central

component of the adipogenic gene regulatory network TL - 3. *Elife* **3**, (2014).

95. Brouzes, E. Single-Cell Analysis, Methods and Protocols TL - 853. *Methods Mol. Biol.* **853**, 105–139 (2012).

96. Consortium, T. *et al.* Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. *Nature* (2018) doi:10.1038/s41586-018-0590-4.

97. Zeisel, A. *et al.* Molecular Architecture of the Mouse Nervous System TL - 174. *Cell* **174** , 999 (2018).

98. Zheng, G. X. Y. *et al.* Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* (2017) doi:10.1038/ncomms14049.

99. Svensson, V. *et al.* Power Analysis of Single Cell RNA - Sequencing Experiments Authors. (2016).

100. Stephenson, W. *et al.* Single-cell RNA-seq of rheumatoid arthritis synovial tissue using low-cost microfluidic instrumentation TL - 9. *Nat. Commun.* **9**, 791 (2018).

101. Gierahn, T. M. *et al.* Seq-Well: portable, low-cost RNA sequencing of single cells at high throughput TL - 14. *Nat. Methods* **14**, 395 (2017).

102. Zilionis, R. *et al.* Single-cell barcoding and sequencing using droplet microfluidics. TL - 12. *Nat. Protoc.* **12**, 44–73 (2017).

103. Han, X. *et al.* Mapping the Mouse Cell Atlas by Microwell-Seq. *Cell* (2018) doi:10.1016/j.cell.2018.02.001.

104. Farrell, J. A. *et al.* Single-cell reconstruction of developmental trajectories during zebrafish embryogenesis. TL - 360. *Science* **360**, (2018).

105. Mayer, C. *et al.* Developmental diversification of cortical inhibitory interneurons TL - 555. *Nature* **555**, 457 (2018).

106. Habib, N. *et al.* Massively parallel single- nucleus RNA-seq with. **14**, (2017).

107. Moon, H.-S. *et al.* Inertial-ordering-assisted droplet microfluidics for high-throughput single-cell RNA-sequencing TL - 18. *Lab Chip* **18**, 775–784 (2018).

108. Chung, M., Núñez, D., Cai, D. & Kurabayashi, K. Deterministic droplet-based co-encapsulation and pairing of microparticles via active sorting and downstream merging TL - 17. *Lab Chip* **17**, 3664–3671 (2017).

109. Klein, J., Fasshauer, M., Klein, H. H., Benito, M. & Kahn, C. R. Novel adipocyte lines from brown fat: a model system for the study of differentiation, energy metabolism, and insulin action. TL - 24. *Bioessays* **24**, 382–388 (2002).

110. Dobin, A. *et al.* STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* (2013) doi:10.1093/bioinformatics/bts635.

111. Gardeux, V., David, F. P. A., Shajkofci, A., Schwalie, P. C. & Deplancke, B. ASAP: a Web-based platform for the analysis and interactive visualization of single-cell RNA-seq data. *Bioinformatics* (2017) doi:10.1093/bioinformatics/btx337.

112. Tang, F. *et al.* mRNA-Seq whole-transcriptome analysis of a single cell TL - 6. *Nat. Methods* **6**, (2009).

113. Zhou, Y. & Yuan, G. Mapping the Mouse Cell Atlas by Microwell-Seq Resource Mapping the Mouse Cell Atlas by Microwell-Seq. *Cell* **172**, 1091-1097.e17 (2018).

114. Rosenberg *et al.* Single-cell profiling of the developing mouse brain and spinal cord

with split-pool barcoding. *Science (80-.).* (2018).

115. Han, X. *et al.* Construction of a human cell landscape at single-cell level. *Nature* (2020) doi:10.1038/s41586-020-2157-4.

116. Saikia, M. *et al.* Simultaneous multiplexed amplicon sequencing and transcriptome profiling in single cells. *Nat. Methods* (2019) doi:10.1038/s41592-018-0259-9.

117. 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.* (2020) doi:10.1038/s41587-019-0372-z.

118. Guide, U. Chromium Single Cell 3' Reagent Kits v3 with FB. *10X Genomics* (2020).

119. Wagner, D. E. *et al.* Single-cell mapping of gene expression landscapes and lineage in the zebrafish embryo. *Science (80-.).* (2018) doi:10.1126/science.aar4362.

120. Serra, D. *et al.* Self-organization and symmetry breaking in intestinal organoid development TL - 569. *Nature* **569**, 66–72 (2019).

121. Grün, D. *et al.* Single-cell messenger RNA sequencing reveals rare intestinal cell types. *Nature* (2015) doi:10.1038/nature14966.

122. Lukonin, I. *et al.* Phenotypic landscape of intestinal organoid regeneration. *Nature* (2020) doi:10.1038/s41586-020-2776-9.

123. Rossi, G., Manfrin, A. & Lutolf, M. P. Progress and potential in organoid research. *Nature Reviews Genetics* (2018) doi:10.1038/s41576-018-0051-9.

124. Gjorevski, N. *et al.* Designer matrices for intestinal stem cell and organoid culture TL - 539. *Nature* **539**, 560 (2016).

125. Brassard, J. A. & Lutolf, M. P. Engineering Stem Cell Self-organization to Build Better Organoids. *Cell Stem Cell* (2019) doi:10.1016/j.stem.2019.05.005.

126. Mustata, R. C. *et al.* Article Identification of Lgr5 -Independent Spheroid-Generating Progenitors of the Mouse Fetal Intestinal Epithelium. 421–432 (2013) doi:10.1016/j.celrep.2013.09.005.

127. Yui, S. *et al.* YAP/TAZ-Dependent Reprogramming of Colonic Epithelium Links ECM Remodeling to Tissue Regeneration. *Cell Stem Cell* (2018) doi:10.1016/j.stem.2017.11.001.

128. Ayyaz, A. *et al.* Single-cell transcriptomes of the regenerating intestine reveal a revival stem cell. *Nature* (2019) doi:10.1038/s41586-019-1154-y.

129. Roulis, M. *et al.* Paracrine orchestration of intestinal tumorigenesis by a mesenchymal niche. *Nature* (2020) doi:10.1038/s41586-020-2166-3.

130. Biočanin, 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. *Lab Chip* **19**, (2019).

131. Unger, M. A. *et al.* Monolithic Microfabricated Valves and Pumps by Multilayer Soft Lithography. **113**, (2014).

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

133. Street, K. *et al.* Slingshot : cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics* **19**, 1–16 (2018).

134. Battich, N. *et al.* Sequencing metabolically labeled transcripts in single cells reveals mRNA turnover strategies TL - 367. *Science (80-.).* **367**, 1151–1156 (2020).

135. Sato, T. *et al.* Single Lgr5 stem cells build crypt–villus structures in vitro without a mesenchymal niche TL - 459. *Nature* **459**, 262 (2009).

136. Haber, A. L. *et al.* Article A single-cell survey of the small intestinal epithelium. *Nat. Publ. Gr.* **551**, 333–339 (2017).

137. Yin, X. *et al.* Niche-independent high-purity cultures of Lgr5+ intestinal stem cells and their progeny TL - 11. *Nat. Methods* **11**, (2013).

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

139. Birchenough, G. M. H., Johansson, M. E. V., Gustafsson, J. K., Bergström, J. H. & Hansson, G. C. New developments in goblet cell mucus secretion and function. *Mucosal Immunology* (2015) doi:10.1038/mi.2015.32.

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

141. Lareau, C. A., Ma, S., Duarte, F. M. & Buenrostro, J. D. Inference and effects of barcode multiplets in droplet-based single-cell assays TL - 11. *Nat. Commun.* **11**, 866 (2020).

142. Lun, A. T. L., Riesenfeld, S., Andrews, T. & Dao, T. P. Distinguishing cells from empty droplets in droplet-based single-cell RNA sequencing data Testing for deviations from the ambient profile. 1–14 (2018).

143. Cheng, Y.-H. *et al.* Hydro-Seq enables contamination-free high-throughput single-cell RNA-sequencing for circulating tumor cells TL - 10. *Nat. Commun.* **10**, 2163 (2019).

144. Zhang, M. *et al.* Highly parallel and efficient single cell mRNA sequencing with paired picoliter chambers. *Nat. Commun.* (2020) doi:10.1038/s41467-020-15765-0.

145. Pollen, A. A. *et al.* Low-coverage single-cell mRNA sequencing reveals cellular heterogeneity and activated signaling pathways in developing cerebral cortex. *Nat. Biotechnol.* (2014) doi:10.1038/nbt.2967.

146. Dura, B. *et al.* Profiling lymphocyte interactions at the single-cell level by microfluidic cell pairing TL - 6. *Nat. Commun.* **6**, 5940 (2015).

147. 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.

148. Haque, A., Engel, J., Teichmann, S. A. & Lönnberg, T. A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications. *Genome Medicine* (2017) doi:10.1186/s13073-017-0467-4.

149. Denisenko, E. *et al.* Systematic assessment of tissue dissociation and storage biases in single-cell and single-nucleus RNA-seq workflows. *Genome Biol.* (2020) doi:10.1186/s13059-020-02048-6.

150. Sato, T. *et al.* Paneth cells constitute the niche for Lgr5 stem cells in intestinal crypts. *Nature* (2011) doi:10.1038/nature09637.

151. Van Es, J. H. *et al.* Notch/γ-secretase inhibition turns proliferative cells in intestinal crypts and adenomas into goblet cells. *Nature* (2005) doi:10.1038/nature03659.

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

153. Hashimshony, T. *et al.* CEL-Seq2 : sensitive highly-multiplexed. *Genome Biol.* 1–7 (2016) doi:10.1186/s13059-016-0938-8.

154. Goldstein, L. D. *et al.* Massively parallel nanowell-based single-cell gene expression profiling. *BMC Genomics* (2017) doi:10.1186/s12864-017-3893-1.

155. Macosko, E. Z. *et al.* Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* (2015) doi:10.1016/j.cell.2015.05.002.

156. Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data Resource Comprehensive Integration of Single-Cell Data. *Cell* **177**, (2019).

157. Saelens, W., Cannoodt, R., Todorov, H. & Saeys, Y. A comparison of single-cell trajectory inference methods. *Nat. Biotechnol.* (2019) doi:10.1038/s41587-019-0071-9.

158. Picelli, S. *et al.* Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. *Genome Res.* **24**, 2033–40 (2014).

159. 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* (2013) doi:10.1038/nmeth.2688.

160. Stoeckius, M. *et al.* Cell ‘hashing’ with barcoded antibodies enables multiplexing and doublet detection for single cell genomics. *bioRxiv* 237693 (2017) doi:10.1101/237693.

161. Honda, Y. *et al.* Guiding the osteogenic fate of mouse and human mesenchymal stem cells through feedback system control. *Sci. Rep.* (2013) doi:10.1038/srep03420.

162. Pittenger, M. F. *et al.* Multilineage potential of adult human mesenchymal stem cells. *Science (80-.).* (1999) doi:10.1126/science.284.5411.143.

163. Haynesworth, S. E., Goshima, J., Goldberg, V. M. & Caplan, A. I. Characterization of cells with osteogenic potential from human marrow. *Bone* (1992) doi:10.1016/8756-3282(92)90364-3.

164. Yianni, V. & Sharpe, P. T. Epigenetic mechanisms driving lineage commitment in mesenchymal stem cells. *Bone* (2020) doi:10.1016/j.bone.2020.115309.

165. Tang, Q. Q., Otto, T. C. & Lane, M. D. Commitment of C3H10T1/2 pluripotent stem cells to the adipocyte lineage. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 9607–9611 (2004).

166. Kolf, C. M., Cho, E. & Tuan, R. S. Mesenchymal stromal cells. Biology of adult mesenchymal stem cells: Regulation of niche, self-renewal and differentiation. *Arthritis Research and Therapy* (2007) doi:10.1186/ar2116.

167. Rauch, A. *et al.* Osteogenesis depends on commissioning of a network of stem cell transcription factors that act as repressors of adipogenesis. *Nat. Genet.* (2019) doi:10.1038/s41588-019-0359-1.

168. Dhaliwal, A., Pelka, S., Gray, D. S. & Moghe, P. V. Engineering Lineage Potency and Plasticity of Stem Cells using Epigenetic Molecules. *Sci. Rep.* (2018) doi:10.1038/s41598-018-34511-7.

169. Fričová, D., Korchak, J. A. & Zubair, A. C. Challenges and translational considerations of mesenchymal stem/stromal cell therapy for Parkinson’s disease. *npj Regenerative*

Medicine (2020) doi:10.1038/s41536-020-00106-y.

170. Rudnicki, M. A. *et al.* MyoD or Myf-5 is required for the formation of skeletal muscle. *Cell* (1993) doi:10.1016/0092-8674(93)90621-V.

171. Cao, J. *et al.* The single-cell transcriptional landscape of mammalian organogenesis. *Nature* 1–7 (2019) doi:10.1038/s41586-019-0969-x.

172. Alda-Catalinas, C. *et al.* A Single-Cell Transcriptomics CRISPR-Activation Screen Identifies Epigenetic Regulators of the Zygotic Genome Activation Program. *Cell Syst.* (2020) doi:10.1016/j.cels.2020.06.004.

173. Xie, S., Duan, J., Li, B., Zhou, P. & Hon, G. C. Multiplexed Engineering and Analysis of Combinatorial Enhancer Activity in Single Cells TL - 66. *Mol. Cell* **66**, 285–29900000 (2017).

174. Replogle, J. M. *et al.* Combinatorial single-cell CRISPR screens by direct guide RNA capture and targeted sequencing. *Nat. Biotechnol.* (2020) doi:10.1038/s41587-020-0470-y.

175. Parekh, U. *et al.* Mapping Cellular Reprogramming via Pooled Overexpression Screens with Paired Fitness and Single-Cell RNA-Sequencing Readout. *Cell Syst.* (2018) doi:10.1016/j.cels.2018.10.008.

176. Xie, S., Cooley, A., Armendariz, D., Zhou, P. & Hon, G. C. Frequent sgRNA-barcode recombination in single-cell perturbation assays. *PLoS One* (2018) doi:10.1371/journal.pone.0198635.

177. Rosen, E. D. *et al.* C/EBPα induces adipogenesis through PPARγ: A unified pathway. *Genes Dev.* **16**, 22–26 (2002).

178. Brun, R. P. *et al.* Differential activation of adipogenesis by multiple PPAR isoforms. *Genes Dev.* (1996) doi:10.1101/gad.10.8.974.

179. Franzén, O., Gan, L. M. & Björkegren, J. L. M. PanglaoDB: A web server for exploration of mouse and human single-cell RNA sequencing data. *Database* (2019) doi:10.1093/database/baz046.

180. Zhao, M. L. *et al.* Molecular Competition in G1 Controls When Cells Simultaneously Commit to Terminally Differentiate and Exit the Cell Cycle. (2020) doi:10.1016/j.celrep.2020.107769.

181. Beekum, O. Van, Fleskens, V. & Kalkhoven, E. Posttranslational modifications of PPAR-γ: Fine-tuning the metabolic master regulator. *Obesity* (2009) doi:10.1038/oby.2008.473.

182. Chen, L., Yu, J. H., Lu, Z. H. & Zhang, W. E2F2 induction in related to cell proliferation and poor prognosis in non-small cell lung carcinoma. *Int. J. Clin. Exp. Pathol.* (2015).

183. Huang, R. *et al.* MYCN and MYC regulate tumor proliferation and tumorigenesis directly through BMI1 in human neuroblastomas. *FASEB J.* (2011) doi:10.1096/fj.11-185033.

184. Zhu, J., Kwan, K. M. & Mackem, S. Putative oncogene Brachyury (T) is essential to specify cell fate but dispensable for notochord progenitor proliferation and EMT. *Proc. Natl. Acad. Sci. U. S. A.* (2016) doi:10.1073/pnas.1601252113.

185. Zare, H., Shooshtari, P., Gupta, A. & Brinkman, R. R. Data reduction for spectral clustering to analyze high throughput flow cytometry data. *BMC Bioinformatics* (2010) doi:10.1186/1471-2105-11-403.

186. Lun, A. T. L., McCarthy, D. J. & Marioni, J. C. A step-by-step workflow for low-level analysis of single-cell RNA-seq data with Bioconductor. *F1000Research* (2016) doi:10.12688/f1000research.9501.2.

187. Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data Resource Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902.e21 (2019).

188. Haghverdi, L., Lun, A. T. L., Morgan, M. D. & Marioni, J. C. Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. *Nat. Biotechnol.* (2018) doi:10.1038/nbt.4091.

189. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* (2009) doi:10.1093/bioinformatics/btp616.

190. Soneson, C. & Robinson, M. D. Bias, robustness and scalability in single-cell differential expression analysis. *Nat. Methods* (2018) doi:10.1038/nmeth.4612.

191. Sergushichev, A. A. An algorithm for fast preranked gene set enrichment analysis using cumulative statistic calculation. *bioRxiv* (2016).

192. Serra, D., Mayr, U., Boni, A., Lukonin, I. & Rempfler, M. Self-organization and symmetry breaking in intestinal organoid development. 1–27.

193. Goetz, J. J. & Trimarchi, J. M. Transcriptome sequencing of single cells with Smart-Seq. *Nat. Biotechnol.* **30**, 763–765 (2012).

194. Abate, A. R., Chen, C.-H., Agresti, J. J. & Weitz, D. A. Beating Poisson encapsulation statistics using close-packed ordering TL - 9. *Lab Chip* **9**, 2628–2631 (2009).

195. Delley, C. L. & Abate, A. R. Modular barcode beads for microfluidic single cell genomics. *bioRxiv* (2020) doi:10.1101/2020.09.10.292326.

196. Hagemann-Jensen, M. *et al.* Single-cell RNA counting at allele- and isoform-resolution using Smart-seq3. *bioRxiv* 817924 (2019) doi:10.1101/817924.

197. Bagnoli, J. W. *et al.* Sensitive and powerful single-cell RNA sequencing using mcSCRB-seq TL - 9. *Nat. Commun.* **9**, 2937 (2018).

198. Lareau, C. A. *et al.* Droplet-based combinatorial indexing for massive-scale single-cell chromatin accessibility. *Nat. Biotechnol.* (2019) doi:10.1038/s41587-019-0147-6.

199. Minn, K. T. *et al.* High-resolution transcriptional and morphogenetic profiling of cells from micropatterned human embryonic stem cell gastruloid cultures. *bioRxiv* (2020) doi:10.1101/2020.01.22.915777.

200. Chen, Q. *et al.* A temporal requirement for Hippo signaling in mammary gland differentiation, growth, and tumorigenesis. *Genes Dev.* (2014) doi:10.1101/gad.233676.113.

201. Johansson, K. A. *et al.* Temporal Control of Neurogenin3 Activity in Pancreas Progenitors Reveals Competence Windows for the Generation of Different Endocrine Cell Types. *Dev. Cell* (2007) doi:10.1016/j.devcel.2007.02.010.

202. Ji, Z. *et al.* Genome-scale identification of transcription factors that mediate an inflammatory network during breast cellular transformation. *Nat. Commun.* (2018) doi:10.1038/s41467-018-04406-2.

203. Ren, X. *et al.* Reconstruction of cell spatial organization from single-cell RNA sequencing data based on ligand-receptor mediated self-assembly. *Cell Res.* (2020) doi:10.1038/s41422-020-0353-2.

204. Cabello-Aguilar, S. *et al.* SingleCellSignalR: inference of intercellular networks from single-cell transcriptomics. *Nucleic Acids Res.* (2020) doi:10.1093/nar/gkaa183.

205. Stickels, R. R. *et al.* Highly sensitive spatial transcriptomics at near-cellular resolution with Slide-seqV2. *Nat. Biotechnol.* (2020) doi:10.1038/s41587-020-0739-1.

206. Raj, A., van den Bogaard, P., Rifkin, S. A., van Oudenaarden, A. & Tyagi, S. Imaging individual mRNA molecules using multiple singly labeled probes. *Nat. Methods* (2008) doi:10.1038/nmeth.1253.

207. Suter, D. M. *et al.* Mammalian genes are transcribed with widely different bursting kinetics. *Science (80-.).* (2011) doi:10.1126/science.1198817.

208. Brunner, A.-D. *et al.* Ultra-high sensitivity mass spectrometry quantifies single-cell proteome changes upon perturbation. *bioRxiv* (2020).

List of abbreviations

BAT – Brown Adipose Tissue
BMP4 – Bone Morphogenic Factor 4
CBC Cells – Crypt Base Columnar Cells
cDNA – Complementary DNA
ChIP – Chromatin Immunoprecipitation
CRISPR – Clustered Regularly Interspaced Short Palindromic Repeats
DAPI – 4′,6-Diamidino-2-Phenylindole
DISCO - Deterministic mRNA-Capture Bead And Cell Co-encapsulation
DNA – Deoxyribonucleic Acid
EDTA – Ethylenediaminetetraacetic Acid
ENCODE – The Encyclopedia of DNA Elements
ESC – Embryonic Stem Cells
ExoI – Exonuclease I
FACS - Fluorescence-Activated Cell Sorting
FANTOM – Functional Annotation Of The Mouse/Mammalian Genome
FBS – Fetal Bovine Serum
GFP – Green fluorescent protein
gRNA – Guide RNA
HA-tag – Human Influenza Hemagglutinin
HEK – Human Embryonic Kidney Cells
hiPSC – Human-Induced Pluripotent Stem Cells
LARRY – Lineage And RNA Recovery
MITOMI – Mechanically Induced Trapping Of Molecular Interactions
mRNA – Messenger RNA
MSC – Mesenchymal Stem Cells
Opti-MEM – Opti Reduced Serum Medium
ORF – Open Reading Frame
PBS – Phosphate-Buffered Saline
PC – Principal Component
PCA – Principal Component Analysis
PCR – Polymerase Chain Reaction
PDMS – Polydimethylsiloxane
PFA – Paraformaldehyde
polyA RNA – Polyadenylated RNA

PSI – Pounds Per Square Inch
qPCR – Quantitave Polymerase Chain Reaction
RNA – Ribonucleic Acid
RNA-seq – Ribonucleic Acid Sequencing
ROI – Region Of Interest
rRNA – Ribosomal Ribonucleic Acid
RT – Reverse Transcription
scATAC-seq – Single-Cell Assay for Transposase-Accessible Chromatin
SCRB-seq – Single-Cell Barcoding and Sequencing
scRNA-seq – Single cell RNA sequencing
SD – Standard Deviation
SELEX – Systematic Evolution Of Ligands By Exponential Enrichment
smFISH – Single Molecule Fluorescent In Situ Hybrdization
SSC – Saline-Sodium Citrate
STAMP – Single Transcriptome Attached to Micro Particle
TE-SDS – Tris EDTA SDS
TE-TW – Tris EDTA Tween 20
TF – Transcription Factor
tSNE – t-Distributed Stochastic Neighbor Embedding
UMAP – Uniform Manifold Approximation and Projection
WAT – White Adipose Tissue

MARJAN BIOČANIN

Address: Avenue de Morges 74, 1004 Lausanne

Contact mail: marjan.biocanin@epfl.ch

HI, NICE TO MEET YOU

I'm a highly motivated and versatile life scientist with experience in **molecular biology, high-throughput technologies, scientific communication, and graphic design** looking for opportunities to solve biomedical challenges to **help bringing more sustainable healthcare solutions worldwide**.

PERSONALITY STRENGTHS

Organized

Communicative

Reliable

Detail-oriented

Analytical thinker

Team player

Proactive

EDUCATION

PhD in **Biotechnology and Bioengineering**, EPFL, IBI, SV, Lausanne, Switzerland (2016 - 2021)

MSc in **Biology**, University of Belgrade, Belgrade, Serbia (2014 - 2015)

BSc in **Biology**, University of Belgrade, Belgrade, Serbia (2010 - 2014)

TRAININGS

RESAL Module 1: Introductory Course in Laboratory Animal Science (RESAL 1/903; 2016)

Swiss Institute of Bioinformatics: Machine Learning for Bioinformatics and Computational Biology (2016)

Swiss Institute of Bioinformatics: Single Cell Sequencing (2016)

Swiss Institute of Bioinformatics: High Performance Computing in Life Science (2015)

EXPERIENCE

APR 2016 -

Laboratory of Systems Biology and Genetics, EPFL, Lausanne, Switzerland

RESPONSIBILITY:

Researcher and doctoral assistant responsible for **development of novel single-cell transcriptomics methods** and methods for **stem cell engineering, data analysis, microfluidic chip fabrication.**

SEPT 2020 - OCT 2020

Drug and Device Product Development and Regulation in Europe and the U.S., UCSD and EPFL

RESPONSIBILITY:

Technical writer responsible for **analysis and document preparation** through case studies to meet the regulatory needs prior to **marketing of drugs and medical devices**.

JULY 2019 -

Infoprisons, Lausanne, Switzerland

RESPONSIBILITY:

Graphic and digital designer responsible for visual content of trimester issued journal discussing prison system and sanction laws in Switzerland.

AUG 2018 - JAN 2020

E3 ambassador, EPFL, Lausanne, Switzerland

RESPONSIBILITY:

EPFL Excellence in Engineering ambassador for Serbia responsible for **external communication** with Faculty heads about school's activities and recruitment.

FEB 2017 - FEB 2019

BioScience Network Lausanne, Switzerland

RESPONSIBILITY:

Communications and content creator responsible for **communication, design, social media, and organization of events and projects** for researchers and entrepreneurs.

EVENTS AND PROJECTS

Fame Lab, Suisse Romande finals (2018, 2019) - responsible within a team for **design, communication and event** logistics.

Open science in Practice summer school at EPFL (2017) - responsible within a team for **communication** with participants and presenters, **marketing**, and **public outreach**.

Exposure: Science film hackathon (2017, 2018) - responsible within a team for **communication and event booklet design**.

SOFTWARE

Specialist: ImageJ

Adobe cloud: Illustrator, Photoshop, After Effects, InDesign, XD, Premiere

Programming: R, bash (basic), Python (basic)

Management: Slack, Trello, Microsoft Office suite, Coggle, SLIMS

INTERNSHIPS

Division of Genomic Technologies, Riken, Yokohama, Japan (2014)

Inserm U1001, Paris, France (2012)

Max Planck Institute for Developmental Biology, Tübingen, Germany (2011)

Institute for molecular biology and genetics, Belgrade, Serbia (2011)

AWARDS

Award for the contribution to the international recognition of the Faculty of Biology, Faculty of Biology, University of Belgrade (2013)

1st place for research work in the field of Biophysics, XVI Annual Youth BIOS Olympiad (2011)

Student annual award from the City of Belgrade (2009 and 2010)

“Dositeja” award for achievements in science, granted by the Fund for Young Talents, Ministry of Youth and Sports of the Republic of Serbia (2010)

LANGUAGES

English - full professional proficiency

French - basic

German - basic

Serbian - mother tongue

PERSONAL INFORMATION

DOB: 07. 06. 1991.

Swiss work permit: B

Citizenship: Serbian

PERSONAL INTERESTS

Street photography

Trumpet playing

Graphic design

Krav maga

SELECTED PUBLICATIONS (5+)

Deterministic scRNA-seq of individual intestinal organoids reveals new subtypes and coexisting distinct stem cell pools, J. Bues, **M. Biočanin**, J. Pezoldt, R. Dainese, A. Chrisnandy, S. Rezakhani, W. Saelens, R. Gupta, J. Russeil, Y. Saeys, E. Amstad, M. Claassen, M. Lutolf, B. Deplancke, (2020), bioRxiv

Multimodal and multisensory coding in the Drosophila larval peripheral gustatory center, L. Maier, **M. Biočanin**, J. Bues, F. Meyenhofer, C.Brunet, J. Y. Kwon, B. Deplancke, S. G. Sprecher (2020), bioRxiv

Simplified Drop-seq workflow with minimized bead loss using a bead capture and processing microfluidic chip, **M. Biočanin**, J. Bues, R. Dainese, E. Amstad, B. Deplancke (2019), Lab on a Chip

Cross-talk between emulsion drops: how are hydrophilic reagents transported across oil phases? G. Etienne, A. Vian, **M. Biočanin**, B. Deplancke, E. Amstad, (2018), Lab on a Chip

Genotypic and Phenotypic Characterization of Stenotrophomonas maltophilia Strains from a Pediatric Tertiary Care Hospital in Serbia, H. Madi, J. Lukić, Z. Vasiljević, **M. Biočanin**, M. Kojić, B. Jovčić, J. Lozo (2016), PloS One

PATENTS

Device for high throughput single-cell studies, B. Deplancke, J. Bues, R. Dainese, **M. Biočanin**, App number 16333297 (2019)

SELECTED PRESENTATIONS (5+)

Deterministic scRNA-seq of individual intestinal organoids reveals new subtypes and coexisting distinct stem cell pools, Single Cell Biology, Cambridge, United Kingdom (2020)

Dissecting adipose cell lineage development using single cell transcriptomics, Swiss Stem Cell Network Annual Meeting, Lausanne, Switzerland (2017)

TEACHING ACTIVITIES

EPFL, Life Sciences faculty, **teaching assistant** to bachelor students in Life Sciences engineering program (2016 - 2018) and **supervision** on individual internship programmes

Petnica Science Center - Serbian Youth Science Center, **teaching assistant** in biology program for young talented high-school students (2011 - 2015)

CERTIFICATES

Axelios: PRINCE2® Foundation Certificate in **Project Management** (GR656179487MB; 2020)

Coursera: **Drug Development Product Management** Specialization (2020)

University of San Diego and EPFL: Drug and Device **Product Development and Regulation** in Europe and the U.S. (2020)

DataCamp: **DataScientist** with R Track (2019)

ONLINE PRESENCE

Scientific Publications

Creative Portfolio

Professional Achievements

*click on the icon to see more