Dissertation
submitted to the
Combined Faculty of Natural Sciences and Mathematics
of the Ruperto Carola University Heidelberg, Germany
for the degree of
Doctor of Natural Sciences

presented by

Seah Yu Fen, Samantha BA
born in Singapore
Oral examination: January 14th 2021

Establishing highly multiplexed microfluidic screening of protein-antibody interactions

Referees: Dr. Oliver Stegle
Prof. Dr. Oliver Fackler

Abstract

Antibodies play an increasingly important role in modern medicine, with monoclonal antibody sales expected to reach nearly USD 200 billion by 2024. High-throughput antibody screening is possible with phage display technologies, however, these are based on antibody binding, which does not necessarily correlate with the phenotypic effects of antibodies on target cells. Traditional functional antibody screening can also be performed with hybridoma cell technology, but as individual hybridoma cells have to be grown up into colonies and tested, this process is both time-consuming and expensive, and only a mere few thousand clones can be obtained and screened.

Droplet microfluidics has been utilised for the screening of individual antibody-secreting cells, as the small droplet volumes enable the accumulation of antibodies to a functional concentration within hours. Droplet microfluidics is also the basis of emulsion-based single-cell transcriptomic assays, where the co-encapsulation of barcoded entities with cells enables the labelling of all mRNA from an individual cell with the same barcode, permitting high-throughput single-cell transcriptomic analyses.

In this thesis, we aimed to evaluate the feasibility of an antibody screening technology that would combine these two applications of droplet microfluidics, and to set up the individual components necessary for such a technology. This technology would enable the study of the effects of single antibody-secreting cells on the transcriptomes of single target cells of interest, in order to identify antibodies of interest.

This requires the co-encapsulation of two different cell types. As this process is governed by Poisson distribution, most droplets cannot reach the desired cell occupancy. To overcome this, we have optimised a picoinjection workflow to selectively inject lysis buffer into droplets with desired cell occupancies, in order to maximise the utility of the sequencing results in our eventual screens.

In addition, we have identified appropriate model systems in which induced transcriptomic changes can be identified at a single-cell level, which can thus be used for further proof of concept and optimisation experiments. We have also evaluated the feasibility of detecting perturbations from the transcriptomic data of a small number of cells treated with drugs, as the identification of antibody hits in our future screens would require the identification of single perturbed transcriptomes in a background of untreated transcriptomes.

Furthermore, as our antibody screening technology would require a means of antibody sequencing, we have developed a Drop-seq-compatible antibody

sequencing methodology that enables the identification and elucidation of heavy and light chain antibody variable region sequences. We apply this methodology to sequence mixtures of four different hybridoma cell lines mixed at different ratios and to sequence a diverse population of antibody-secreting cells, for which we have no prior knowledge of the heavy and light chain variable regions present.

These individual components pave the way for the development of a microfluidic antibody screening pipeline that employs single-cell transcriptomics. By studying gene expression changes as a proxy for the global phenotypes of target cells, the effects of different antibodies on various different targets can be simultaneously monitored, permitting highly multiplexed screening campaigns.

Zusammenfassung

Antikörper spielen in der modernen Medizin eine immer wichtigere Rolle. Bis 2024 wird der jährliche Umsatz von monoklonalen Antikörpern einen Wert von voraussichtlich bis zu 200 Mrd. USD erreichen. Mit Hilfe von Phagen-Display-Technologien sind Hochdurchsatz-Antikörper-Screenings möglich. Diese basieren jedoch auf Antikörperbindungen und repräsentieren nicht notwendigerweise auch funktionale Effekte wie z.B. (ant)agonistische Wirkungen auf Zellrezeptoren.

Auf der anderen Seite kann die Hybridom-Technik für ein funktionelles Antikörper-Screening eingesetzt werden. Dieser Prozess ist allerdings sowohl zeitaufwändig als auch kostspielig, da einzelne Hybridomzellen zu Kolonien herangezüchtet werden müssen, und nur ein Paar tausend Klone können erhalten und untersucht werden.

Für das Screening einzelner Antikörper-sekretierender Zellen wurde auch die Tröpfchenmikrofluidik eingesetzt, da die kleinen Tröpfchenvolumina die Akkumulation von Antikörpern bis zu einer funktionellen Konzentration innerhalb von Stunden ermöglichen. Die Tröpfchenmikrofluidik ist auch die Grundlage für transkriptomische Einzelzell-Assays auf Emulsionsbasis, bei denen die Co-Einkapselung von Barcode-Entitäten mit diesen Zellen, zur Markierung aller mRNAs aus ein und derselben Zelle mit dem gleichen Barcode führt. Dies ermöglicht transkriptomische Einzelzell-Analysen im Hochdurchsatz.

Das Ziel dieser Dissertation war es, die Realisierbarkeit einer Antikörper-Screening-Technologie, die diese beiden Anwendungen der Tröpfchenmikrofluidik kombinieren würde, zu evaluieren und die notwendigen Einzelkomponenten aufzubauen. Diese Technologie würde die Wirkung individueller Antikörper-sekretierender Zellen auf das Transkriptom einzelner Zielzellen von Interesse untersuchen, um die funktionellen Antikörper zu identifizieren.

Diese Technologie erfordert die gemeinsame Einkapselung von zwei verschiedenen Zelltypen. Da diese Einkapselung durch eine Poisson-Verteilung bestimmt ist, weisen viele Tröpfchen nicht die gewünschte Zellbelegung auf. Um dieses Problem zu lösen, haben wir einen Pico Injektions-Workflow optimiert, bei dem selektiv nur Lysepuffer in Tröpfchen mit der gewünschten Zellbelegungen injiziert wird. Dadurch wird die Nützlichkeit der Sequenzierungsdaten für unsere vorgesehenen Screens maximiert.

In diesem Zusammenhang wurden geeignete Modellsysteme etabliert, in denen induzierte transkriptomische Veränderungen auf Einzelzellenebene analysiert

werden können. Darüber hinaus wurde untersucht, ob die Wirkung von Medikamenten aus den transkriptomischen Daten einer kleinen Anzahl von behandelten Zellen abgeleitet werden kann, da die Selektion von funktionalen Antikörpern in unseren zukünftigen Screens die Identifizierung einzelner veränderter Transkriptome vor dem Hintergrund unbehandelter Transkriptome erfordert.

Zusätzlich bedarf unsere Antikörper-Screening-Technologie einer geeigneten Methode, um die korrekten Sequenzen der Antikörper zu erhalten. Aus diesem Grund wurde eine Drop-seq-kompatible Antikörper-Sequenzierungsmethode entwickelt, welche die Identifizierung von Sequenzen variabler Regionen der schweren und leichten Ketten von Antikörpern ermöglicht. Wir wenden diese Methode auf Mischungen von vier verschiedenen Hybridomzelllinien in unterschiedlichen Mischverhältnissen an, sowie auch auf die Sequenzierung einer diversen Population von Antikörper-sekretierenden Zellen, bei denen zuvor keine Vorkenntnisse über die variablen Regionen der schweren und leichten Ketten herrschten.

Zusammengenommen ebnen diese Methoden den Weg für die Entwicklung einer mikrofluidischen Antikörper-Screening-Pipeline mit Hilfe der Einzelzell-Transkriptomik. Durch die Untersuchung von Genexpressionsänderungen als Proxy für den globalen Phänotyp von Zielzellen können die Auswirkungen verschiedener Antikörper auf verschiedene Zielstrukturen gleichzeitig überwacht werden, was wiederum hochmultiplexierte Screening-Kampagnen ermöglicht.

Acknowledgements

The thesis is the culmination of four years of hard work and it would not have been possible without the help and support of my mentors, colleagues, family and friends.

Firstly, I would like to thank Prof. Dr. Christoph Merten for the opportunity to work in his lab at EMBL Heidelberg. His scientific input has been indispensable for the work described in this thesis and for my scientific growth. Our interactions have also contributed to my professional and personal growth. I am grateful for the time spent in his lab at EMBL, where I have learnt so much more than just microfluidics. A big thanks goes to the members of my Thesis Advisory Committee, Prof. Dr. Lars Steinmetz, Dr. Oliver Stegle and Prof. Dr. Oliver Fackler, for their scientific input and feedback over the four years. I would also like to express my gratitude to Dr. Oliver Stegle and Prof. Dr. Oliver Fackler, for being the examiners of this thesis, and to Prof. Dr. Lars Steinmetz and Prof. Dr. Stefan Wölfl for agreeing to be on my defence committee.

I am very grateful to all past and present members of the Merten Group that I have had the honour of working with. Your scientific input, advice and help was very helpful at all stages of my PhD, and I am grateful to you all for the support provided over all these years. I would like to especially thank Dr. Hongxing Hu, whom I have collaborated closely with over the past four years, and Dr. Xiaoli Ma, who has contributed to the bioinformatic analyses in this thesis. An exceptional thanks goes to Martine Ballinger, for the scientific and personal support provided, especially during the tougher times. Extra thanks go to Anna Mei, a 100% determined colleague whom I have had the absolute honour of collaborating with, learning from and mentoring.

The Genome Biology Computational Support team has been indispensable in realising this project. I would like to thank all members, especially Dr. Charles Girardot and Jelle Scholtalbers, who analysed the transcriptomic and antibody sequencing data and provided invaluable feedback at many stages of the project. The EMBL Genomic Core Facility has also been of immense help. Special thanks go to Dr. Vladimir Benes, Dr. Jonathan Landry, Bianka Baying, Nayara Azevedo and Dinko Pavlinic for their help and scientific input. I would also like to thank the EMBL Flow Cytometry Facility, in particular Dr. Malte Paulsen and Dr. Diana Ordoñez, for their scientific help and for the training provided. I am also grateful to Klaus Schmitt and the team at the EMBL Mouse Facility for their help with mouse immunisations. The EMBL Graduate Office has also been invaluable. I would like to express my utmost gratitude to all past and current members for their support, especially to Dr. Monika Lachner, who has provided an incredible amount of administrative and personal support.

I am thankful to all our collaborators who have contributed to this project. I would like to thank Prof. Michael Ryckelynck from the University of Strasbourg, who shared the picoinjection technology with us. I would also like to thank Prof. Patricia McDonald for sharing the GLP-1R and Exendin-4 constructs and reagents with us.

I am grateful to the Agency for Science, Technology and Research (A*STAR), Singapore for the funding they have provided, under the National Science Scholarship (NSS-PhD), and to various individuals from A*STAR who have supported me both scientifically and administratively.

Many thanks to Dr. Jervis Thevathasan for introducing me to LaTeX and for the formatting of this thesis. Many thanks also go to Szilárd Library for their help with literature access and to Alexandra Elbakyan for her contributions to open science. I would also like to thank Anna Mei, Christian Müller, Leonie Kolmar and Samuel Pažický for proofreading the Zusammenfassung.

The many friends that I have made in EMBL have made this journey much more enjoyable. I would like to thank you all for your scientific and personal help. In particular, I would like to thank Samuel Pažický for his support and for the amazing times over the past four years. Also, a huge shoutout to my Singaporean friends in EMBL (Dr. Jervis Thevathasan, Zhang Hui Ting, Lim Jia Le and Yong Luo Yan) and to the Durians (Dr. Esther Kim, Dr. Hugo Samano-Sanchez, Carlos Rojas Cordova and Paul Gerard Sanchez) for the great company and support when the going got tough. Thank you to all other friends, near and far, for your support, especially those who took the time to check in, despite the distance.

Not forgetting the various activist communities and societies in EMBL that I have had the honour of being a part of - Feminism in Science Book Club, greenEMBL - thank you for being the change that we need and for making EMBL a better place. I am also immensely grateful to the various global scientific communities and initiatives that I have been a part of, specifically preLights and the various initiatives that grew from the eLife Community Ambassador Programme (Environmental Sustainability in Science, Improving Conferences, Intersectionality in Science and the Intersectionality Book Club). I have learnt and grown so much from my interactions with you and my PhD journey has benefited greatly from my participation in these initiatives.

Last but not least, I would like to thank my family for their unwavering love and support. I am immensely blessed to have a large family located all around the world - thank you for supporting me both from near and far. Special thanks goes to my parents for their unconditional love over all these years - they have helped me become who I am today. Thank you!

Personal Contributions

The work presented in this thesis were parts of collaborative efforts. My contributions are as follows:

Chapter 3: Preparation of a complex hybridoma library - I contributed to experimental design, carried out experimental work and data analysis.

Part I: A single-cell transcriptomic method to map cell-cell interactions

Chapter 4: A workflow for the transcriptomic analysis of cell-cell interactions - I contributed to experimental design.

Chapter 5: Identification of a suitable model system - I contributed to experimental design and carried out experimental work and data analysis.

Chapter 6: Bioinformatic identification of positive events - I contributed to experimental design, carried out experimental work and gave input on data analysis.

Part II: A single-cell antibody sequencing workflow

Chapter 9: Antibody sequencing of predefined mixtures of hybridoma cells and **Chapter 10: Antibody sequencing of a complex hybridoma mixture** - I contributed to experimental design and carried out experimental work and data analysis.

Contents

List of Abbreviations

1α,25-(OH)$_2$D$_3$	1,25-dihydroxyvitamin D3
2°	Secondary
3D	Three-dimensional
Ab	Antibody
ACE-1	Angiotensin-converting enzyme
APC	Antigen-presenting cell
ASC	Antibody-secreting cell
BD	Becton, Dickinson and Company
BCR	B-cell receptor
BFP	Blue fluorescence protein
bp	Base pairs
BSA	Bovine serum albumin
C	Celsius
cDNA	Complementary deoxyribonucleic acid
CD	Cluster of differentiation
CDR	Complementarity-determining regions
CEL-seq	Cell expression by linear amplification and sequencing
CFSE	Carboxyfluorescein succinimidyl ester
CHO	Chinese Hamster Ovary cells
CO_2	Carbon dioxide
CITE-seq	Cellular indexing of transcriptomes and epitopes by sequencing
cm	Centimeters
CRISPR	Clustered regularly interspaced short palindromic repeats
CROP-seq	CRISPR droplet sequencing
dA	deoxyadenine
Da	Daltons
DAPI	4,6-diamidino-2-phenylindole
DART-seq	Droplet-assisted RNA targeting by single-cell sequencing
DC	Dendritic cell
ddPCR	Droplet digital polymerase chain reaction
DMEM	Dulbecco's Modified Eagle's Medium
DMF	Dimethylformamide

DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dNTP	deoxyribose nucleoside triphosphate
dPCR	Digital polymerase chain reaction
dT	deoxythymine
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
ELISA	Enzyme-linked immunosorbent assay
Ex-4	Exendin-4
FACS	Fluorescence-activated cell sorting
FADS	Fluorescence-activated droplet sorting
FBS	Fetal bovine serum
FcR	Fc receptor
FcRn	Neonatal Fc receptor
FDA	Food and Drug Administration
FMAT	Fluorometric microvolume assay technology
FRET	Förster resonance energy transfer
g	Gravity-force
GFP	Green fluorescent protein
GLP1	Glucagon-like peptide-1
GLP1R	Glucagon-like peptide-1 receptor
GPCR	G protein-coupled receptor
h	Hours
HAT	Hypoxanthine-aminopterin-thymidine
HCl	Hydrochloric acid
HEK	Human embryonic kidney cells
HER2	Human epidermal growth factor receptor 2
HGPRT	Hypoxanthine-guanine phosphoribosyltransferase
HTO	Hashtag oligo
IFN	Interferon
IgG	Immunoglobulin G
IM	Intramuscular
IMDM	Iscove's Modified Dulbecco's Medium
ITO	Indium tin oxide
kDa	Kilodaltons
kHz	Kilohertz
kNN	k-nearest neighbour
LBP	Lipopolysaccharide-binding protein
LPS	Lipopolysaccharides
M	Molar
mAb	Monoclonal antibody
MACS	Magnetic associated cell sorting

mESC	Mouse embryonic stem cell
MHV	Murine hepatitis virus
min	Minutes
ml	Milliliters
mm	Millimeters
MOI	Multiplicity of infection
Mosiac-Seq	Mosiac single-cell analysis by indexed CRISPR sequencing
mRNA	messenger ribonucleic acid
MUSIC	model-based understanding of single-cell CRISPR screening
NaCl	Sodium chloride
ng	Nanograms
NGS	Next generation sequencing
NK	Natural Killer cells
nm	Nanometers
PBS	Phosphate buffered saline
PCA	Principal component analysis
PCR	Polymerase chain reaction
PDMS	Polydimethylsiloxane
PEG	Polyethylene glycol
PFO	1H,1H,2H,2H-Perfluorooctanol
PFPE	Perfluoropolyether
PMA	Phorbol 12-Myristate 13-Acetate
PMT	Photomultiplier tube
Pol II	RNA Polymerase II
pRb	Retinoblastoma protein
PTFE	Polytetrafluoroethylene
PS-1	Picosurf-1
QC	Quality control
qPCR	Quantitative polymerase chain reaction
RAGE-seq	Repertoire and Gene Expression by Sequencing
RNA	Ribonucleic acid
RNA-seq	RNA sequencing
RPMI-1640	Roswell Park Memorial Institute 1640 media
rRNA	ribosomal ribonucleic acid
RSV	Respiratory syncytial virus
RT	Reverse transcription
s	Seconds
SARS-CoV-2	Severe acute respiratory syndrome coronavirus 2
SC	Subcutaneous
SCDE	Single-cell differential expression
scFv	Single-chain variable fragment
sci-RNA-seq	single-cell combinatorial indexing RNA sequencing

SDS	Sodium dodecyl sulfate
sgRNA	Single guide RNA
SMART	Switching mechanism at the 5′ end of the RNA transcript
SPR	Surface plasmon resonance
ssDNA	single-stranded deoxyribonucleic acid
STAMPs	Single-cell Transcriptomes Attached to MicroParticles
t	Time
TAP-seq	Targeted Perturb-seq
TBS	Tris buffered saline
TCR	T-cell receptor
TNFα	Tumour necrosis factor alpha
t-SNE	t-distributed stochastic neighbor embedding
UMAP	Uniform Manifold Approximation and Projection
UMI	Unique molecular identifier
UV	Ultraviolet
VSV-G	Vesicular stomatitis virus glycoprotein
WT	Wild Type
µg	Micrograms
µl	Microliters
µm	Micrometers
µM	Micromolar
ZGA	Zygotic genome activation

For those who have lost, or who will lose their lives, homes or livelihoods to climate change

1 Introduction

Microfluidics encompasses various technologies and approaches in which small amounts of fluids are manipulated and analysed. This is typically done within micron-scale structures with defined geometries (i.e. channels). The first systems were made of silicon (Manz et al., 1990), while the platforms used today are mostly PDMS- (Whitesides, 2006), paper- or fiber-based (Hosseini et al., 2017). The small volumes used in microfluidic systems, usually in the femtolitre to microlitre scale, enable higher throughput with the same reagent use and make it viable to analyse biological samples that are only available in small volumes. In addition, the small sample volumes mean a reduction in dilutions, which make the detection of low-abundance molecules possible.

Early microfluidic devices were manufactured from silicon and glass using clean room techniques that were already established for the semiconductor industry (Sackmann et al., 2014). However, these materials have their limitations: silicon is opaque to visible light and both glass and silicon devices are complicated to fabricate (Convery and Gadegaard, 2019).

The use of polydimethylsiloxane (PDMS) together with lithography revolutionised the field of microfluidics (Whitesides, 2006), by making microfluidics accessible to the wider research community. A combination of high-resolution printing and photolithography is used to produce master moulds displaying a positive relief of the desired channel structures (Qin et al., 1996). Chips are produced by pouring PDMS over the structures before the cured PDMS is removed from the moulds and sealed by covalent bonding to glass or PDMS, frequently via plasma oxidation (Duffy et al., 1998). The production of PDMS chips from master moulds is cheap and simple, and can be carried out in the absence of specialised clean room equipment. A detailed workflow of the fabrication of microfluidic devices can be found in the Materials and Methods (Chapter 13).

Additionally, the elasticity of PDMS permits not only the easy removal from moulds for feature replication, but also has huge advantages for valving and actuation (Quake and Scherer, 2000; Unger et al., 2000). PDMS is also both transparent (Piruska et al., 2005) and effectively biocompatible, although unpolymerized crosslinker can be present in cured PDMS, and is toxic if not

removed. The biocompatibility is due in part to its high permeability to gas (Lamberti et al., 2014), which enables the long-term culture and imaging of cells and organisms (Choudhury et al., 2012; Huberts et al., 2013), although this also means that evaporation can occur during experiments (Berthier et al., 2008).

1.1 Physics of microfluidics

The small dimensions present in microfluidics result in fluids having different properties when compared to larger fluidic systems. Predicting which form of fluidic flow will dominate in any given system is done by calculating Reynold's number (Re), a dimensionless number outlining the ratio between inertial and viscous forces,

$$Re = \frac{\rho U_0 L_0}{\mu} \tag{1.1}$$

where ρ is the fluid density, U_0 is the fluid velocity, L_0 is the length scale of the system and μ is the fluid viscosity.

For small values of Reynold's number ($Re < 10$), flow is laminar, while for large values of Reynold's number ($Re > 1000$), flow is turbulent. In microfluidic systems, L_0, the length scale of the system, is generally in the scale of micrometres and the velocities (U_0) are typically in the range of 1 µm/s – 1 cm/s. Therefore, the Re values for microfluidic devices are usually between 10^{-6} and 10^{1} (Squires and Quake, 2005). This suggests that viscous forces will prevail, and that two or more miscible liquids flowing parallel in a microfluidic channel will follow a laminar flow regime, where they will not mix except by diffusion.

A second dimensionless number, the Péclet number (Pe), is crucial in the modelling of the relative importance of convection and diffusion,

$$Pe = \frac{U_0 w}{D} \tag{1.2}$$

where U_0 is the fluid velocity, w is the width of the channel and D is the diffusion coefficient of the molecule of interest. The number of channel widths required for full mixing varies proportionately with Pe. For example, in a typical T-junction where two liquids are injected to flow alongside each other, a small protein flowing in a fluid through a 100 µm channel at 100 µm/s would require 250 channel widths to mix. In microfluidic devices, channel length correlates with time, so this would be approximately 2.5 cm, or 4 minutes (Squires and Quake, 2005).

By optimising system dimensions and experimental parameters while bearing the Péclet number in mind, on-chip gradients can be generated and reliably maintained.

1.2 Single-phase microfluidics in biology

The most basic single-phase microfluidic system is a microfluidic chamber filled with a single fluid (Fig. 1.1A). This is a powerful tool for the culture and imaging of single cells (Skelley et al., 2009), organoids (Karzbrun et al., 2018), embryos (Chung et al., 2011) and multicellular organisms (Chokshi et al., 2009).

Furthermore, microfluidics enables the generation of different conditions within a single chamber, enabling the analysis of cellular and organismal responses to spatial and temporal changes in stimuli. Utilising laminar flow, Lucchetta *et al.* subjected the different halves of single *Drosophila melanogaster* embryos to different temperatures for different periods of time and examined changes in development (Lucchetta et al., 2005).

The utilisation of PDMS structures with specific geometries permits the study of cell motility and transportability. Liu *et al.* developed a microfluidic chip that is capable of sorting cells based on cell size and the ability of cells to travel through small constrictions. This was made possible by the use of a special channel geometry involving tilted microposts (Liu et al., 2015b). When used in combination with stable chemokine gradients, complex chemotactic behaviours can be analysed. This has been illustrated by Jain *et al.* in their study of T-lymphocyte exploration patterns. They utilised a chip containing a maze structure, which mimicked a tissue-like environment, together with a chemokine gradient induced across the chip. The tightly defined geometrical and chemical features enabled the precise quantification of the changes in cell movement induced by cell activation and by the presence of chemoattractant gradients (Jain et al., 2015).

Other geometries, namely the "Organ-on-a-Chip" or "Body-on-a-Chip" systems, utilise microfluidics to reconstitute more complex systems. These model crucial processes within or between specific organs or the body's dynamic response to various drugs. This was first demonstrated for a three chamber Microscale Cell Culture Analog (μCCA) device that includes both lung and liver cells (Sin et al., 2004), while similar systems were subsequently utilised for elucidating and evaluating drug mixtures as potential cancer treatments (Tatosian and Shuler, 2009).

1.2.1 Valves in microfluidics

Valving and actuation systems enable compartmentalisation within single-phase microfluidic systems (Fig. 1.1C). Pneumatic microvalves, first developed in the Quake lab (Unger et al., 2000), are the most widely used. These microvalves consist of two layers separated by a PDMS membrane (Fig. 1.1C, grey), with the flow layer containing a liquid phase. The application of pneumatic pressure via the control layer results in a downward deflection of the membrane, closing the fluidic channel and interrupting flow through the flow layer.

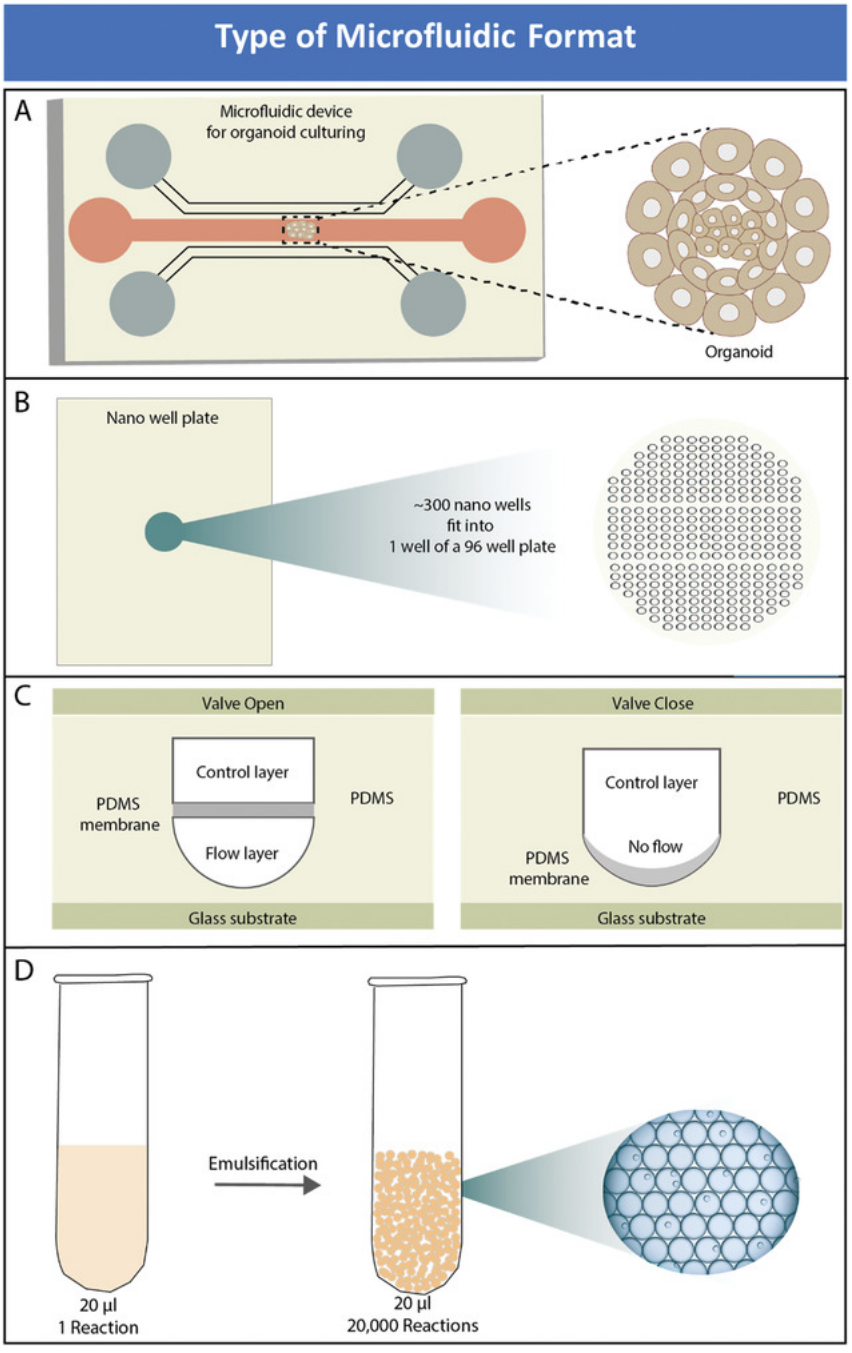


Figure 1.1: Different microfluidic formats utilised in biology. **A.** Single-phase microfluidic chambers can be used for cell culture. **B.** Nanowells can be used for the large-scale compartmentalisation of cells. **C.** Pneumatic valves are composed of two layers. The application of pressure results in downward deflection of the membrane, closing the fluidic channel. **D.** Emulsification enables the efficient generation of a large number of individual compartments. Figure reprinted from Mathur et al., 2019 under the terms of the Creative Commons Attribution License (CC-BY 4.0).

These have enabled the development of complex microfluidic systems that are capable of delivering multiple inputs and monitoring multiple outputs from cells. This has been demonstrated by Junkit *et al.*, where single macrophages were studied upon stimulation with bacterial lipopolysaccharides (LPS) (Junkin et al., 2016). The microfluidic system contained 16 reagent inputs, and antibody-functionalised beads were utilised to study cytokine levels. Cells were also subjected to on-chip imaging, immunostained, or isolated for cell culture or gene expression analysis, illustrating the complexity that can be achieved with valve-based systems.

1.3 Nanowells in biology

Any discussion of the role that microfluidics plays in biology would not be complete without the mention of nanowells, which are wells present in a microfluidic chip, with volumes in the nanolitre range. Even though the utilisation of nanowells does not necessarily entail the usage of microfluidics, they are similar to microfluidic technologies in their device production and the role they play in biological analyses.

Like other microfluidic devices, nanowells can be produced from PDMS (Fig. 1.1B), and can be easily scaled up, thus enabling massive parallelisation. Depending on their size, these nanowells fulfil a plethora of functions, ranging from the generation and study of spheroids (Kim et al., 2012; Ruppen et al., 2015) to the high-throughput analysis of single cells (Han et al., 2012). Nanowells have been utilised by the Love group to study T-lymphocyte activation at a single-cell level. Single T-lymphocytes were trapped and the secretion of cytokines was measured every two hours after stimulation (Han et al., 2012), illustrating how these systems permit the high-throughput, multiplexed and dynamic study of complex cellular processes on a single-cell level.

1.4 Two-phase microfluidics

Two-phase microfluidics utilises two immiscible liquids to generate discrete units of one liquid (droplets or plugs) in a continuous phase of the other liquid. In this thesis, we refer largely to aqueous droplets in a continuous oil phase, as this is more useful for our biological applications. Both the inverse, oil droplets in a continuous aqueous phase, as well as double emulsions, both water-in-oil-in-water (W/O/W) and oil-in-water-in-oil (O/W/O), are also feasible.

As millions of surfactant-stabilised droplets can be generated at kilohertz (kHz) frequencies, the scale of compartmentalisation possible with two-phase microfluidics is several orders of magnitude higher than in the other microfluidic systems outlined previously (Fig. 1.1D).

1.4.1 Droplet generation

The interaction of two immiscible phases is defined by various physical properties, with the most important being surface tension (γ). As these properties remain consistent within each microfluidic system, microfluidic systems are deterministic, facilitating the reliable and reproducible generation of monodisperse emulsions (Squires and Quake, 2005).

Various channel geometries can be used for monodisperse droplet generation with two immiscible phases (Fig. 1.2). The main geometries utilised are T-junctions (Thorsen et al., 2001), flow-focusing junctions (Anna et al., 2003) and co-flow systems (Umbanhowar et al., 2000).

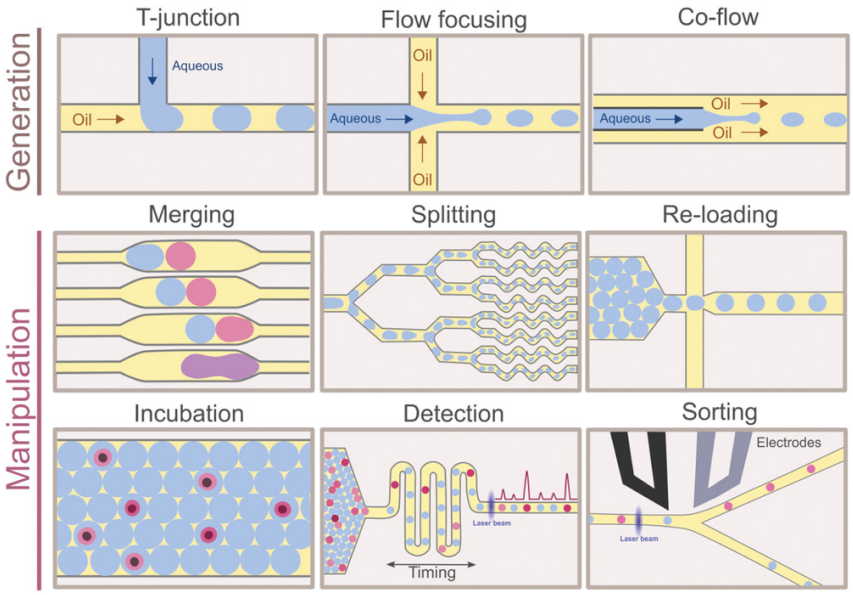


Figure 1.2: Droplet generation and manipulation modules. Droplet generation is carried out by primarily three different geometries: T-junctions, flow-focusing, and co-flow modules. Subsequently, droplets can be manipulated with a variety of manipulation modules. Droplets can be merged with other droplets, split into smaller droplets, re-loaded into a different microfluidic device and incubated on chip. In addition, the fluorescence signals in droplets can be detected, and droplets of interest can be sorted. Figure reprinted from Matuła et al., 2020 under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License (CC-BY-NC-ND 4.0).

For T-junctions, droplets are generated by injecting the aqueous stream orthogonally to the oil phase. The presence of interfacial tension between the two fluids causes competition between interfacial tension and viscous stresses, resulting in instability and droplet formation. Changing the relative flow rates and the channel shapes influences the droplet size and morphology (Thorsen et al., 2001).

Flow-focusing junctions generate droplets by adding oil from two sides orthogonal to the aqueous phase, together with a constriction just downstream which opens into a wide channel (Anna et al., 2003). This draws the fluid into a narrow stream which then breaks up by Rayleigh-Plateau instability, which is the

phenomenon where a jet of liquid breaks into multiple droplets (Rayleigh, 1879). This is driven by the fact that surface tension forces tend towards minimising the surface area between the two phases. Small sinusoidal perturbations in the jet of liquid grow when the radius of the curvature of the waves is larger than the radius of the cylinder of the inner liquid, eventually leading to rupturing of the pinched areas to form droplets.

In co-flow systems, the tip of a capillary is used to inject an aqueous stream into a stream of oil. Drops form at the tip of the capilliary and detach once the drag caused by the co-flowing liquid exceeds the interfacial tension (Umbanhowar et al., 2000).

To generate stable emulsions that can be both stored and used as bioreactors, the interfacial tension between the aqueous and oil phases must be reduced, to prevent droplet fusion upon contact. This is achieved by the addition of amphiphilic molecules known as surfactants (surface active agents) to the continuous phase. These surfactants must fulfil three important functions, namely stability, control of exchange between droplets and biocompatibility (Baret, 2012). Many applications today utilise fluorinated oils as these are biocompatible (Giaever and Keese, 1983), permit gas exchange and most organic compounds are insoluble in them (Lemal, 2004). Correspondingly, various fluorosurfactants have been developed for use in droplet microfluidics. Those made of perfluorinated poly-ethers (PFPE) coupled to polyethylene-glycol (PEG) are highly biocompatible and have been used widely (Holtze et al., 2008), although alternatives continue to be developed (Chowdhury et al., 2019; Wagner et al., 2015). These fluorinated oils and fluorosurfactants ensures that cells and other biological entities can be encapsulated into droplets and that cells can be kept alive over several days (Clausell-Tormos et al., 2008).

In addition to droplet generation, there exists various modules for the manipulation of droplets, to mimic manipulations which are crucial for biological experimentation on a larger scale, such as pipetting (Fig. 1.2). Droplets can be split into smaller droplets by the use of tree-branching geometries (Fig. 1.2, Splitting) (Link et al., 2004). For processes requiring incubation steps, droplets can be incubated on chip (Fig. 1.2, Incubation) or off chip prior to re-injection into a microfluidic system (Fig. 1.2, Re-loading). The detection of fluorescent signals in droplets can be carried out by the use of lasers and photo-multiplier tubes (PMTs) (Fig. 1.2, Detection) (Baret et al., 2009) and droplets can be merged with other droplets (Fig. 1.2, Merging) (Bremond et al., 2008) or sorted (Fig. 1.2, Sorting) (Niu et al., 2007), either simultaneously with or independently of droplet detection.

1.4.2 Droplet detection

The selective high-throughput processing of generated droplets is only possible in the presence of a high-throughput means by which droplets can be analysed.

Fluorescence-based analysis of droplets fits this mold, due to the low timescales (sub-milliseconds) involved in fluorescence detection by PMTs. A laser beam of the appropriate wavelength is focused on the channels, and emitted light of the right wavelength is quantified by the use of band-pass filters to ensure that only light of the appropriate wavelengths reach the associated PMT. Multiple lasers and PMTs can be combined within a single droplet detection system, enabling the simultaneous analysis of multiple fluorescence parameters, and thus the development of complex high-throughput microfluidic assays where droplet contents can be rapidly and reliably identified and analysed.

This has been utilised for combinatorial drug screening, where an optical setup comprising of three different excitation lasers (375, 488 and 561 nm) and readouts at three different wavelengths (450, 521 and >580 nm) permitted the use of highly multiplexed assays in the blue, green and red spectrum respectively within a single experiment (Eduati et al., 2018). In this setup, the authors utilised plugs, which are large non-surfactant-stabilised droplets that occupy the the entire channel or tubing, and these were separated by mineral oil droplets. This experimental setup keeps plugs in the same order as they were generated.

The authors produced experimental plugs and used a rhodamine 110 (green fluorescent dye) conjugated substrate of Caspase-3, an early marker of apoptosis, and Alexa Fluor 594 (orange fluorescent dye) to track apoptosis and the addition of cells to the assay respectively. Furthermore, a set of plugs containing binary concentrations (high/low) of the blue fluorescent dye cascade blue were produced before each set of experimental plugs, to both separate groups of plugs and to identify plugs for each experimental condition. These cascade blue plugs encoded numbers in binary code, for example, 2 plugs with high-low blue fluorescence encoded the binary number "10", which is the decimal number "2". As a number was assigned to each experimental condition, this permitted the retrospective elucidation of the identity of individual experimental plugs. The plugs were produced and stored in PTFE tubing and were analysed by flushing the plugs through a detection module capable of detecting readouts at three different wavelengths, to reveal information on droplet identity, cell apoptosis and cell loading.

1.4.3 Droplet sorting

Active droplet sorting is carried out to enrich for specific droplets of interest. In the case of fluorescence-activated droplet sorting (FADS), fluorescence readouts are used to trigger sorting decisions, in a manner similar to fluorescence-activated cell sorting (FACS). Droplets are analysed upstream of the sorting junction. By default, all droplets are directed to the waste channel. However, when a droplet fulfils certain fluorescence requirements, an electric field is applied which deflects the droplet into the collection channel via dielectrophoresis (Ahn et al., 2006). The detection and sorting of entire droplets allows for secreted products to be detected

and permits assays involving multiple cells that would not be possible with FACS (Baret et al., 2009).

1.4.4 Droplet fusion and picoinjection

Droplet fusion is crucial for adding of reagents to droplets, which is essential for biological workflows in which reagents must be added to samples in a sequential manner.

Droplet fusion can be carried out in a passive manner, where droplets are induced to come into contact and fuse by the use of different channel geometries, such as an expansion of channel width to induce a destabilisation of droplet interfaces (Bremond et al., 2008) or the use of pillar elements to slow and deform droplets to enable droplet fusion (Niu et al., 2008). Most passive droplet fusion modules utilise droplets with no or little surfactant, such that there remains significant interfacial tension between the phases, enabling droplet fusion based on surface tension phenomena.

In contrast, electro-coalescence permits the fusion of surfactant-stabilised droplets. The droplets of interest are subjected to an electric field, which triggers a dipole in both droplets to be fused. When the droplets are near each other, their induced dipoles align, leading to an attractive force between them, and thus to droplet fusion (Chabert et al., 2005). The fusion of two different droplet types has been demonstrated at kHz frequencies (Mazutis et al., 2009).

Reagents can also be added to pre-existing droplets by the use of picoinjection (Fig. 1.3). Surfactant-stabilised droplets are re-injected into a chip and spaced with oil. Droplets pass by a picoinjector nozzle, through which the reagent of interest is added. Picoinjection only takes place when the electrodes are activated, creating an electric field which destabilises the surfactant film surrounding droplets and permits reagents from the nozzle to be injected into the passing droplets (Abate et al., 2010). The amount of reagent injected can be varied by varying the injection velocity, which is determined by the injection pressure at the nozzle, and by varying the injection time, which is inversely related to the oil flow rate. In addition, picoinjection can be carried out for all droplets or selectively triggered for a subset of droplets, making it a useful tool for the selective processing of droplets. More details can be found in Sections 4.3 and 13.1.6 of this thesis.

Both active droplet fusion via electro-coalescence and picoinjection are presently used to add reagents to droplets. As electro-coalescence requires the synchronisation of multiple droplet streams to enable reliable droplet pairing, it can be tricky to set up, especially if re-injected emulsions are involved (Eastburn et al., 2013a). The volumes to be added may also affect the choice of method; laboratories working with both electro-coalescence and picoinjection typically say that picoinjection works

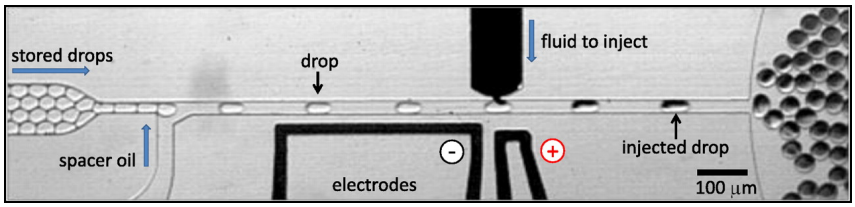


Figure 1.3: Microfluidic device with a droplet spacer and a picoinjector. Droplets are spaced by the addition of oil. The reagent is added by the picoinjector, which is a pressurised channel through which the reagent is injected. The application of an electric field by the electrodes destabilised the surfactant film surrounding the droplet, permitting the reagents to merge with the droplet passing the picoinjector. Figure reprinted from Abate et al., 2010 with permission from PNAS.

better for the addition of small volumes, while electro-coalescence works better for adding larger volumes (Personal Communication, Ryckelynck, 2020).

1.5 Emulsion polymerase chain reaction

1.5.1 Polymerase chain reaction

The indispensable polymerase chain reaction (PCR) technique, which permits the amplification of small amounts of genetic material, was first conceived by Kary Mullis in 1984 (Saiki et al., 1985). The initial PCR method, together with a majority of modern-day PCR methods, depend on thermal cycling, where repeated cycles of heating and cooling permit various temperature-dependent reactions to take place. The reaction requires primers, which are short DNA oligonucleotides that are complementary to sequences flanking the region of interest, and DNA polymerase, an enzyme that synthesises DNA from free nucleotides in the presence of a DNA template. Firstly, denaturation of DNA, where the two strands of DNA are physically separated, is carried out at over 90 °C. Then, the temperature is lowered to around 50 °C to 65 °C, permitting the binding of primers to template DNA, before enzyme-catalysed DNA replication occurs at 72 °C, where DNA polymerase synthesises new DNA from free nucleotides, based on the DNA template present (Quan et al., 2018). The utilisation of heat-stable polymerases, such as the Taq polymerase from *Thermus aquaticus*, removes the need to continually replenish polymerase, which had to be done previously when heat-sensitive polymerases were used. As the DNA generated in each cycle is itself used as template in subsequent cycles, exponential amplification of the DNA takes place.

Apart from the amplification of genetic material, PCR has also been adapted for the quantification of the amount of genetic material present within samples, via quantitative PCR (qPCR) (Arya et al., 2005). As the amplification of DNA is exponential, there is a quantitative relationship between the amount of starting template and the amount of PCR product at any particular cycle, but this plateaus off as reagents are used up or due to the accumulation of inhibitory products (Arya

et al., 2005). Therefore, such quantification is done by monitoring the amplification of a target DNA molecule throughout the PCR, rather than at the end. This is done by coupling a fluorescent readout to the amount of DNA generated, by using either DNA intercalating dyes or hydrolysis-based probes, and by monitoring fluorescence over time.

While qPCR has been widely used to measure genomic DNA and cDNA levels, the resulting data can be variable and problematic, particularly for samples with low levels of nucleic acids and/or variable amounts of chemical and protein contaminants. Under these circumstances, digital PCR has been shown to produce more precise and reproducible data (Taylor et al., 2017).

1.5.2 Digital polymerase chain reaction

Digital polymerase chain reaction (dPCR) is a relatively new method for the absolute quantification of nucleic acids. The technique has been gaining popularity for its precision and sensitivity and is increasingly utilised in both healthcare and environmental analyses.

The first step in dPCR is the subdivision of the reaction into many independent sub-reactions, such that every partition contains most likely one or no template molecules. PCR is then carried out, and the proportion of positive (with template, and thus DNA amplification) and negative (no template and no amplification) partitions is then elucidated. These are digital signals, as opposed to the analog signals obtained during qPCR (Quan et al., 2018; Whale et al., 2016). These digital signals permit precise quantification using Poisson statistics, which is less problematic compared to the calibration curves and variations in reaction efficiencies seen in qPCR quantification (Svec et al., 2015). In addition, the small partition volumes used effectively increase the concentrations of templates within the individual reaction vessels and additionally separates competing templates. This reduces template competition and enhances the detection of rare variants in a background of wild-type templates. This is also expected to increase tolerance to inhibitors, given that targets are concentrated within smaller volumes (Quan et al., 2018).

While dPCR has been carried out in microtubes (Sykes et al., 1992), microtitre plates (Vogelstein and Kinzler, 1999) and non-droplet microfluidic setups, including valves (Unger et al., 2000), nanowells (Matsubara et al., 2004) and chambers (Du et al., 2009; Heyries et al., 2011), droplet microfluidics has been widely adopted for digital PCR, in a format called droplet digital PCR (ddPCR), due to the ease and scalability of droplet generation. Here, isolated microreactors in the form of aqueous droplets are created within an immiscible oil phase.

The chemistry required for dPCR, and thus ddPCR, is broadly similar to that used for qPCR. For both applications, a higher fluorescence should be

present upon increased amplification. Thus, either non-specific fluorescent dyes or sequence-specific DNA probes fused to a fluorescent reporter dye and a quencher are used, with the fluorescent readout in dPCR analysed only after the entire PCR has been carried out. The readout can be obtained via sequentially interrogating droplets with fluorescence excitation sources and detectors (as explained in Section 1.4.2) or via wide-field microscopy. The ddPCR technology is commercially available, for example, as the Bio-Rad QX200^{TM} Droplet Digital^{TM} PCR System.

Like most other droplet microfluidic processes involving fluorescence readouts, the throughput of ddPCR is currently largely limited by the readout. Alternatives, such as converting droplets into cytometry-compatible agarose beads (Zhang et al., 2012) or double emulsions (Zinchenko et al., 2014), could be considered to overcome this shortcoming (Quan et al., 2018).

1.5.3 Emulsion PCR for compartmentalisation

Emulsion PCR has also been utilised in other contexts, specifically where the segregation of different templates is required or beneficial, such as in the amplification of genomic and cDNA libraries. The bulk amplification of such complex mixtures of nucleic acids is usually hindered by two problems, namely that shorter fragments tend to be preferably amplified over larger fragments, and that artifacts can be generated via recombination between homologous regions of DNA (Williams et al., 2006). Therefore, droplet PCR has been utilised for the amplification of complex DNA mixtures (Williams et al., 2006), including for the preparation of complex libraries for molecular evolution and aptamer selection (Shao et al., 2011).

While bead-based barcoding approaches, as exemplified by CytoSeq (Fan et al., 2015), Drop-seq (Macosko et al., 2015) and InDrop (Klein et al., 2015), were developed first and are more widely utilised for molecular barcoding, this can also be carried out with droplet barcoding libraries, where each droplet contains many copies of the same barcode sequence, but different droplets have different barcode sequences (Lan et al., 2016; Lan et al., 2017; Shahi et al., 2017). Emulsion PCR is utilised in the generation of droplet barcoding libraries, where barcode oligonucleotides flanked by constant sequences are encapsulated at a limiting dilution, together with PCR reagents and primers complementary to the constant regions. One primer contains a sequence that enables the fragments to cluster on Illumina flow cells. The droplets are then subject to thermal cycling, enabling the generation of approximately ten million barcode droplets in a few hours. The droplets are subsequently fused with droplets containing templates or cells of interest and further processed to label all DNA fragments from a single droplet with the same barcode sequence. This enables all reads corresponding to a given template or cell to be identified via the barcode. The use of molecular barcoding in single-cell genetic analyses will be further elaborated on in Section 1.6.1.

Emulsion PCR is similarly crucial in the context of fusion, or overlap extension PCR, specifically when there is variability in template sequences and therefore, templates should not be fused indiscriminately.

A key example is the cloning and sequencing of antibody heavy and light chains, which are highly variable within a single organism, and thus within a single sample. As the heavy and light chains are encoded by different genes, the pooling and bulk sequencing of B-lymphocytes inevitably results in a loss of endogenous heavy and light chain pairings. Similarly, overlap extension PCR of the pooled heavy and light chain sequences will produce non-endogenous fusion products.

This problem has been overcome by the use of emulsion PCR, where antibody heavy and light chain pairings can be maintained. This has typically been carried out via workflows involving two rounds of encapsulation. Firstly, single antibody-secreting cells are encapsulated with lysis buffer and poly-dT beads for mRNA capture, such that every poly-dT bead contains genetic material from a single cell. The poly-dT beads are then subjected to a second round of encapsulation at a limiting dilution, such that most droplets maximally contain one bead. Overlap extension PCR is then carried out to fuse the heavy and light chain sequences, in a manner that maintains endogenous pairings. These fragments can then be prepared for sequencing (DeKosky et al., 2015; McDaniel et al., 2016, Sumida et al., 2012), or even expressed as yeast display libraries and used for the screening of antibody binding (Adler et al., 2017a; Adler et al., 2017b). These techniques, and other antibody sequencing technologies, will be further explained in Section 1.8.4.

Similarly, emulsion overlap extension PCR has been utilised for the characterisation of microbial communities, where functional target genes of interest were fused with 16S rRNA genes. Cells were encapsulated together with acrylamide, which was then polymerised, trapping the cells and their genomes in polyacrylamide beads. As the polyacrylamide permits the diffusion of enzymes, primers and buffers, these beads were directly enzymatically treated to enable cell lysis. The beads were then re-encapsulated and subject to overlap extension PCR, with the fused amplicons only present in droplets where the microbial cell contains the functional gene of interest. This technique provides information on both the genes of interest and the host organisms containing those genes, as microbial species can be identified by their 16S rRNA sequences (Spencer et al., 2016).

Overall, the ability of droplet microfluidics to partition PCR reactions has played a central role in advancing quantification via digital PCR and has similarly elevated PCR applications where compartmentalisation is required. The effortlessness by which this has been employed and commercialised reflects the ease in which molecular processes can be adapted to droplet formats. This paves the way for more complex enzymatic reactions, assays and workflows to be conducted in droplets, including, but not limited to, those crucial for single-cell analyses.

1.6 Single-cell analyses

Due to cellular heterogeneity present in eukaryotic populations and systems (Rosenfeld et al., 2005; Weinberger et al., 2005), single-cell analyses are critical in the discovery of mechanisms and details that cannot be detected via bulk analyses. For example, single-cell analyses have enabled the discovery and further analysis of rare cell types in the immune system (Villani et al., 2017). Single-cell analyses are also important in cancer genetics, where genetic changes result in intra-tumour heterogeneity with consequences on treatment outcome (Saadatpour et al., 2015). Furthermore, the recombination events in B- and T-lymphocytes (Tonegawa, 1983) and meiosis in the production of sperm (Wang et al., 2012) create heterogeneous populations, such that their functional characteristics and diversity can only be assessed at the single-cell level.

1.6.1 Molecular barcoding in single-cell genetic analyses

Molecular barcoding in single-cell genetic analyses involves the labelling of genetic material from each single cell with nucleic acid sequences unique to each cell, to enable the mapping of all obtained reads back to their cells of origin. While this strategy has been used in single-cell genomic (Lan et al., 2017) and epigenetic analyses (Lareau et al., 2019), this section will focus on single-cell transcriptomic analyses.

The labelling of all mRNA from each single cell in single-cell transcriptomic analyses is done by the tagging of all mRNA transcripts from each individual cell with cell-specific barcodes, and this process takes place during the reverse transcription of the mRNA into cDNA (Fig. 1.4).

This is made possible by the introduction of oligonucleotides containing sequences permitting amplification and sequencing (Univ), cell barcodes (CL), UMIs and poly-dT sequences (Fig. 1.4). The poly-dT sequence enables the capture of mRNA (Fig. 1.4A) and the subsequent reverse transcription, to generate cDNA fused to the universal sequences, cell barcodes and UMIs (Fig. 1.4B). By introducing the same cell barcode to all cDNA from the same cell, reads can be mapped back to their cells of origin after sequencing. This can be done with a barcoded bead, as illustrated in Figure 1.4 and exemplified by CytoSeq (Fan et al., 2015), Drop-seq (Macosko et al., 2015) and InDrop (Klein et al., 2015), by the use of droplet barcoding libraries (Lan et al., 2016; Lan et al., 2017; Shahi et al., 2017), or by the pre-spotting of barcode-containing oligonucleotides within nanowells (Goldstein et al., 2017).

An alternative means of molecular barcoding involves combinatorial indexing, where cells or nuclei are subject to split-and-pool barcoding, such that the transcriptomes of individual cells or nuclei are uniquely labelled (Cao et al., 2017; Cusanovich et al., 2015). This is exemplified by sci-RNA-seq, where fixed and permeabilised cells are first subdivided into 96- or 384-well plates, with each well

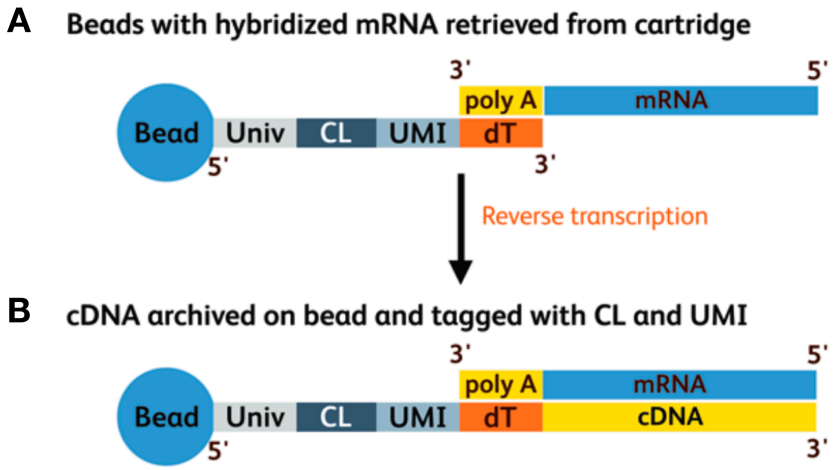


Figure 1.4: An example of beads and oligonucleotides utilised for cell barcoding in single-cell transcriptomic analyses. The oligonucleotides utilised in the BD Rhapsody™ Single-Cell Analysis System are pictured. **A.** Oligonucleotides utilised for single-cell transcriptomic analyses contain sequences permitting amplification and sequencing (Univ = Universal sequence, light grey), a cell barcode (CL = Cell Label, dark grey), unique molecular identifier (UMI, blue-grey) and a poly-dT sequence (orange). mRNA sequences are captured by hybridisation of the poly-dA tails (yellow) to the poly-dT sequences (orange). **B.** Subsequent reverse transcription results in the archiving of cDNA on the bead and the labelling with the cell label and the UMI. Figure reprinted from Shum et al., 2019 with permission from Springer Nature.

receiving a molecular index via *in situ* reverse transcription. Cells are then pooled and redistributed via FACS into 96- or 384-well plates in limiting numbers, before being subjected to second-strand synthesis, transposition with transposon 5 (Tn5) transposase, lysis, and polymerase chain reaction (PCR) amplification, where a second well-specific barcode is introduced. As a large majority of cells pass through a unique combination of wells, individuals cells can be distinguished by the unique barcode combination they receive. The rate of "collisions", where cells receive the same combination of indices, can be finely tuned by regulating the number of cells distributed to each well during the second round of indexing.

The length of the nucleic acid barcode utilised in molecular barcoding naturally affects the possible diversity of the barcode pool. If we assume that all possible nucleic acid sequences of a given length N are utilised, the number of possible barcodes will be equivalent to 4^N. Therefore, a 12-bp long barcode can assume 4^{12} (approximately 10^7) different sequences and a 30-bp barcode can assume 4^{30} (approximately 10^{18}) different sequences (Kebschull and Zador, 2018). The diversity of the barcode pool is evidently also dependent on how many different sequences are accessible; some techniques intentionally limit the diversity of barcodes to simplify subsequent data processing. For example, in 10X Genomics, barcodes are drawn from a pre-determined 'whitelist' of sequences, to simplify subsequent error-correction and read assignment (Zheng et al., 2017). In addition, the number of barcodes accessible in sci-RNA-seq is rather dependent on the number of wells

utilised during the two indexing steps, as opposed to the length of the barcodes, although the barcode lengths must be sufficient to ensure diversity across the wells utilised (Cao et al., 2017).

In order to ensure that cells are uniquely labelled, the labelling of multiple cells with the same barcode must be avoided. For combinatorial indexing strategies, as previously mentioned, this can be regulated by modulating the number of cells distributed to each well during the second round of indexing. For non-combinatorial indexing strategies, this is most easily done by ensuring that the diversity of barcodes is orders of magnitudes larger than the number of cells (Kebschull and Zador, 2018). If we assume that the sampling of barcodes from the pool is random, the required barcode diversity can be easily estimated.

For the labelling of k cells with barcodes that are drawn from a pool of N barcodes, for which $N >> k$, and where every barcode is equally likely to be chosen, the fraction of uniquely labelled cells (F) can be estimated as

$$F \approx 1 - \frac{k}{N} \tag{1.3}$$

Therefore, if the diversity of the barcode pool is 100-fold larger in diversity than the number of labelled cells ($N/k = 100$), there would be 99% unique labelling (Kebschull and Zador, 2018), which is suitable for most single-cell genomic applications.

1.6.2 Nanowells and valve-based single-cell genetic analyses

Both valves and nanowells have been utilised for single-cell genetic analyses, due to their abilities to segregate and isolate single cells for independent processing.

Single cells can be segregated into nanowells and processed independently. Cells are seeded into the wells using a limiting dilution - based on Poisson distribution, a majority of wells would be empty or would contain a single cell. Cellular barcoding takes place via the addition of barcoded beads (Fan et al., 2015; Gierahn et al., 2017) or by the pre-spotting of oligonucleotides containing barcodes within the wells (Goldstein et al., 2017). The former has been commercialised by BD Biosciences as the BD Rhapsody™ Single-Cell Analysis System, while the latter has been commercialised by TakaraBio as the iCELL8 system. In the iCELL8 system, nanowells can be imaged and the dispensing of cells and reagents is highly regulated, overcoming the problems of Poisson distribution and enabling the efficient single-cell transcriptomic profiling of thousands of single cells (Kim et al., 2019).

In contrast, valve-based systems enable the isolation of single cells into fluidically segregated compartments. These are the basis of the commercially available Fluidigm C1 system, which is capable of trapping and barcoding single

cells for over thirty different applications (Lynch and Ramalingam, 2019), including genomic (Fan et al., 2011; Wang et al., 2012), transcriptomic (Streets et al., 2014), epigenomic (Buenrostro et al., 2015) and even combinatorial analyses (Buenrostro et al., 2018; Dey et al., 2015; Macaulay et al., 2015).

1.6.3 Emulsion-based single-cell analyses

As previously mentioned, the formation of water-in-oil droplets creates a large number of individual reaction vessels. By including cells in the aqueous phase, cells can be encapsulated and segregated into individual droplets, which then act as independent reaction vessels. The occupancy of droplets depends on cell concentration and droplet size, and this follows a Poisson distribution (Clausell-Tormos et al., 2008). As droplet size remains consistent for any given combination of flow rates in a droplet generation chip of a specific geometry, cell occupancy can be modified by changing the cell concentration used. In many analyses, the average number of cells per drop (λ) used, which can be calculated as the cell density divided by the droplet volume, would be around 0.1. This minimises the probability of obtaining droplets containing multiple cells ($p < 0.005$), but results in a majority of droplets containing no cells ($p \approx 0.90$) (Shembekar et al., 2016). However, this is inconsequential as high frequencies of droplet production (tens of kHz) nonetheless enable the rapid generation of large numbers of cell-containing droplets. In addition, such loading densities make it highly probable that cell-containing droplets contain exactly a single cell. This makes droplet microfluidics a popular tool to study single cells.

Various genomic (Lan et al., 2017), transcriptomic (Klein et al., 2015; Macosko et al., 2015; Zheng et al., 2017), epigenomic (Lareau et al., 2019; Rotem et al., 2015) and proteomic analyses for membrane proteins (Peterson et al., 2017; Shahi et al., 2017; Stoeckius et al., 2017b) and secreted proteins (El Debs et al., 2012; Eyer et al., 2017; Mazutis et al., 2013) that utilise droplet microfluidics for compartmentalisation of single cells have been developed.

In this section, we will focus on methods for single-cell transcriptomics and single cell proteomics for the detection of membrane proteins. Single-cell droplet microfluidic methods focusing on secreted proteins, specifically antibodies, are explained in Sections 1.9.3 and 1.9.4.

1.6.4 Emulsion-based single-cell transcriptomics

As previously outlined, single-cell transcriptomic analyses were possible with technologies that did not utilise droplet microfluidics. However, the development of two high-throughput droplet microfluidic single-cell transcriptomic technologies, inDrop (Klein et al., 2015) and Drop-seq (Macosko et al., 2015) in 2015, and the subsequent introduction of the 10X Genomics Chromium platform for 3′ Gene

Expression analysis (Zheng et al., 2017) revolutionised the field of single-cell transcriptomics.

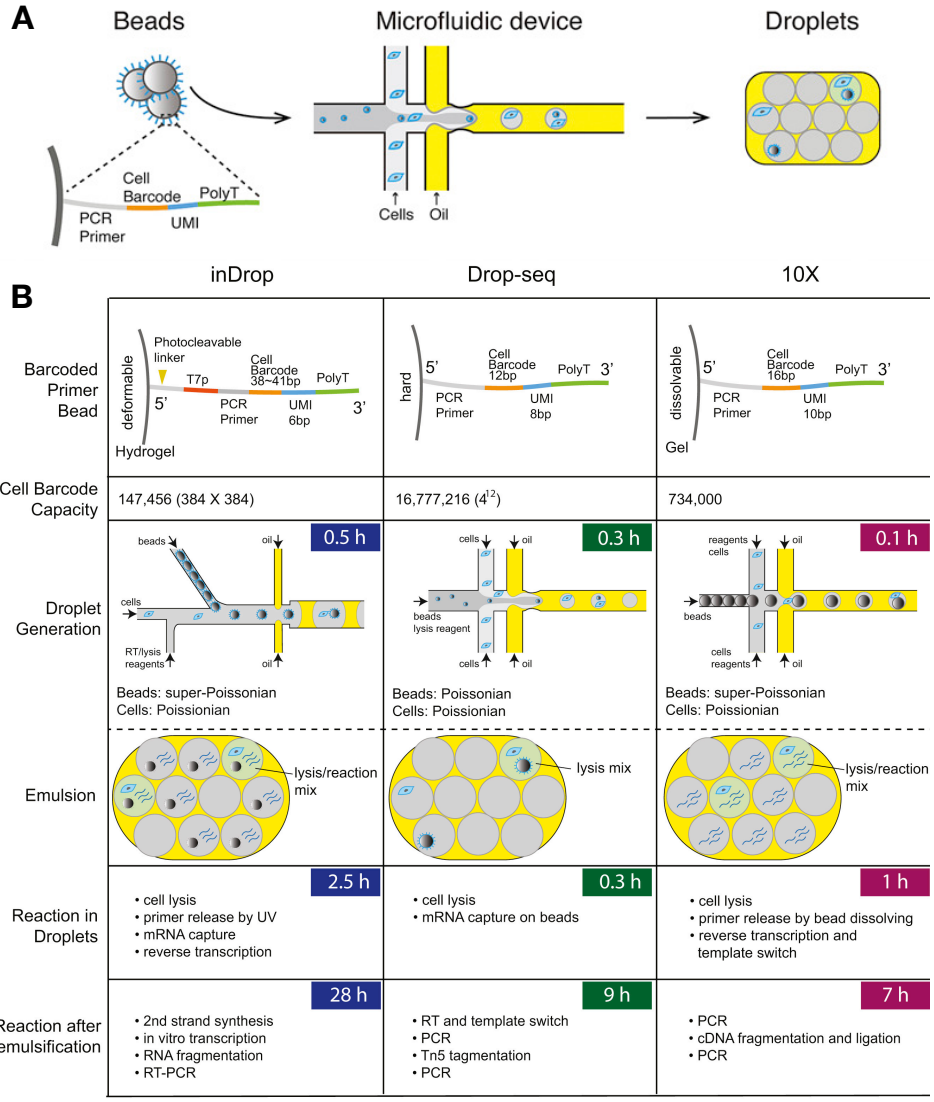


Figure 1.5: Summary and comparison of emulsion-based single-cell transcriptomic technologies. A. Summary of the three emulsion-based single-cell transcriptomic technologies. Beads are functionalised with oligonucleotides containing PCR primer, a cell barcode, UMI and a poly-dT tail. These beads are co-encapsulated with cells. **B.** Comparison of the three emulsion-based single-cell transcriptomic technologies. They differ in the structure and oligonucleotide sequences of the barcoded bead, the cell barcode capacity, droplet generation method and workflow. Figure reprinted from Zhang et al., 2019 with permission from Molecular Cell.

The three technologies utilise similar principles to enable single-cell transcriptomic analyses (Fig. 1.5), and are fairly simple to implement as they merely require a single droplet maker.

For all three technologies, cells are co-encapsulated together with beads that have been functionalised with oligonucleotides consisting of a PCR primer, cell barcode, unique molecular identifier (UMI) and poly-dT tail (Fig. 1.5A). The presence

of the poly-dT tail enables the capture of mRNA. Cellular barcoding is made possible by the inclusion of a cell barcode in the primers. As previously outlined, all primers on the same bead contain the same cell barcode, thus all mRNA from a single cell would be labelled with the same barcode. After sequencing, the sequenced fragments can then be mapped back to the original droplet, and thus to the original cell that they came from.

The unique molecular identifiers (UMIs) are random oligonucleotide sequences of 4 - 15 bp, which differ between different oligonucleotides associated to each bead. This means that reads with the same cell barcode and the same UMI and that come from the same transcript originate from a single mRNA molecule, as opposed to reads that map to the same gene, and contain the same cell barcode but different UMIs, which would instead be derived from unique mRNA transcripts from the same gene and expressed in the same cell (Fig. 1.6). This enables correction for amplification bias and a more accurate quantification of mRNA transcripts (Islam et al., 2014).

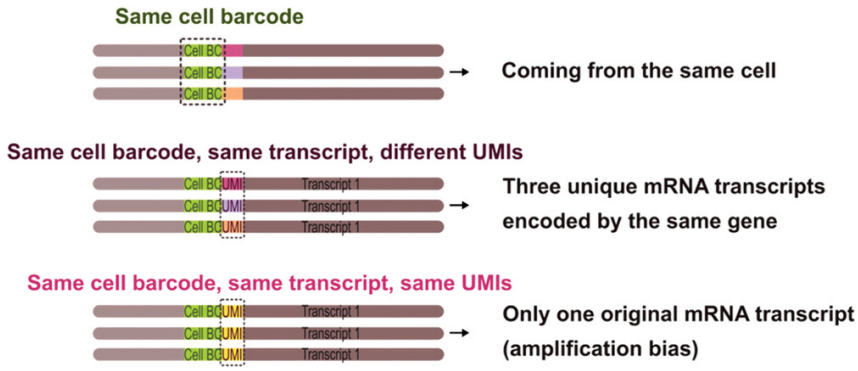


Figure 1.6: Cell barcodes and UMIs alleviate amplification bias and increase the accuracy of mRNA transcript quantification. Reads with the same cell barcode come from the same cell. Reads with the same cell barcode and transcript read, but different UMIs, are derived from unique mRNA transcripts, while reads with the same cell barcode, same transcript read and same UMI are ultimately derived from the same mRNA transcript. Figure reprinted from Matuła et al., 2020 under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License (CC-BY-NC-ND 4.0).

A key difference between the different single-cell transcriptomic technologies is the type of barcoding bead used (Fig. 1.5B). In Drop-seq, 30 µm beads made of a hard resin are used, while in inDrop and 10X Genomics, 60 - 70 µm deformable hydrogel and gel beads are used respectively. The different physical properties affect the encapsulation efficiency – the Drop-seq beads are encapsulated with a Poisson distribution, which means that a mere 2 - 4% of cells are captured (Macosko et al., 2015). In contrast, the beads used for inDrop and 10X Genomics are deformable and can be closely packed, which means that they can be loaded with a super Poisson distribution, where 60 - 90% of cells are barcoded in inDrop (Klein et al., 2015). The beads also have different chemical properties that affect their handling. The oligonucleotides attached to the hydrogel beads utilised for inDrop contain a

photo-cleavable spacer and are thus released by exposure to UV (Klein et al., 2015), while the beads utilised in 10X Genomics dissolve upon the co-encapsulation of cells, due to the simultaneous addition of a reducing agent, which reduces disulfide linkages present within the beads (Bent et al., 2020; Zheng et al., 2017). This means that while the beads utilised for Drop-seq can be utilised for multiple rounds of library preparation, this is not the case for inDrop or 10X Genomics.

The differences in fabrication also results in different levels of bead size variation, where it was noted that the beads used in Drop-seq were less uniform in size compared to those of 10X Genomics and inDrop (Zhang et al., 2019). Zhang *et al.* hypothesise that this has consequences for the determination of which barcodes are associated with healthy cells. The total number of raw reads is frequently used to differentiate between healthy cells and dead cells or empty droplets. While a distinct difference is noted between barcodes from healthy cells and other droplets in 10X and InDrop, such a difference is not seen for Drop-seq (Zhang et al., 2019).

Various processing steps differ between the technologies. Both Drop-seq and the 10X Genomics technology utilise the Moloney murine leukemia virus (MMLV) reverse transcriptase, which adds a few cytosine nucleotides at the 3′ end of the cDNA, but only when the polymerase reaches the 3′ end of the cDNA and not when cDNA synthesis is prematurely terminated. The SMART (Switching mechanism at the 5′ end of the RNA transcript) technology permits the addition of a PCR handle at the 3′ end with the use of a template switch oligonucleotide (TSO) (Zhu et al., 2001).

In contrast, inDrop utilises CEL-seq (Cell Expression by Linear amplification and Sequencing) (Hashimshony et al., 2012), where *in vitro* transcription (IVT) linear amplification followed by RT-PCR is used to amplify the barcoded cDNAs. This potentially reduces PCR amplification bias and noise (Ziegenhain et al., 2017).

Other differences exist based on which enzymatic processes take place in droplets versus in bulk (Fig. 1.5B). In Drop-seq, reverse transcription and template switching occur in bulk, while in inDrop, reverse transcription is carried out in droplets. In the 10X Genomics system, reverse transcription, template switching and transcript extension occur in droplets. Carrying out reverse transcription in droplets is necessary in inDrop and for the 10X Genomics technology, due to primer release and dissolving of the beads, respectively.

Reverse transcription in bulk, as is carried out for Drop-seq, could increase the risk of cross-contamination, although this could be minimised with optimisation of the washing steps (Biočanin et al., 2019). Yet, reverse transcription in bulk could reduce the time at which the reverse transcriptase has to remain at room temperature before reverse transcription and also eliminate the problem of RT inhibition in droplets by mammalian cell lysates (Eastburn et al., 2013b; Mary et al., 2011).

There are also differences in the length of the cell barcodes and UMIs (Fig. 1.6). The differing structures and lengths of the barcodes affect the theoretical cell barcode library size, and thus the maximum cell capacity that can be processed in a single run. Additionally, the barcodes used in InDrop and in 10X Genomics are generated from a known 'whitelist' of sequences, while the barcodes in Drop-seq are completely random. This has consequences on the actual diversity generated, where the diversity in InDrop and 10X Genomics is predefined by the number of different sequences in the whitelists, rather than a mathematical function of the number of bases (as mentioned in Section 1.6.1). This also has consequences on data analysis, where the use of a 'whitelist' permits simplification of error-correction, while the synthesis strategy in Drop-seq provides no prior information to aid error-correction or read assignment (Tambe and Pachter, 2019). There are also differences in the cell barcode error rates for the different systems, which are likely to arise during DNA synthesis. It appears that 10X beads have the lowest error rates, and thus are of the best quality (Zhang et al., 2019).

In addition, errors in UMIs, including point mutations and a bias towards certain nucleotides or sequences introduced during UMI generation, result in a reduction in effective UMI space and in incorrectly reduced molecular counts, specifically amongst highly expressed genes. Longer UMIs, such as those seen in the 10X Genomics platform, could mitigate these effects (Petukhov et al., 2018).

Overall, the 10X Genomics Single Cell 3' Gene Expression technology appears to outperform the other two technologies in sensitivity, precision and cell barcode quality (Ding et al., 2019; Zhang et al., 2019). Additionally, the 10X Genomics platforms require the smallest time investment, although they cost about twice as much per cell compared to Drop-Seq and inDrop (Ding et al., 2019). In addition, as the 10X Genomics platform is commercial, it is constantly subject to changes and improvements that continue to improve its efficiency and throughput. For example, alternative library construction methods, such as assessing 5' gene expression via the Single Cell 5' Gene Expression kit, have been introduced. Naturally, the commercial and black box nature of the 10X Genomics systems makes them less amenable for technological development and modification within academic environments.

1.6.5 Analysis of emulsion-based single-cell RNA-seq data

A key challenge of single-cell RNA-seq data analysis is that the data is inherently noisy, due to the susceptibility of single-cell RNA-seq to various sources of technical and biological variation. Technical variation originates from variation in capture efficiency, amplification bias and variation in sequencing efficiency (Bacher and Kendziorski, 2016; Hwang et al., 2018), while biological variation arises from differences in cell-cycle stage or cell size in different cells, which are not routinely of interest for most experiments (Bacher and Kendziorski, 2016). Such biological

variation is typically not detected in bulk RNA-seq due to averaging over large cell numbers.

Another challenge for single-cell RNA-seq is the large number of "dropouts", which generally refers to observed zeros, where a given gene in a given cell has no UMIs or reads mapping to it (Lähnemann et al., 2020). While some of these events may be due to genes that are biologically not expressed, other events may be attributed to technical limitations, such as systematic mRNA degradation during cell lysis, or events that occur by chance, such as the failure to capture transcripts that are expressed at low levels (Bacher and Kendziorski, 2016; Hicks et al., 2018). It has been estimated that only approximately 10% to 20% of all transcripts are captured and sequenced (Kolodziejczyk et al., 2015). This is problematic, as it has been shown that the proportion of genes with a zero expression level is a big source of cell-cell variation and, together with batch effects, can cause false discoveries (Hicks et al., 2018). While the degree of sparsity varies between single-cell RNA-seq techniques, emulsion-based techniques are particularly susceptible to such problems (Ding et al., 2019), as these high-throughput techniques are typically sequenced in a shallow manner (Eraslan et al., 2019; Kashima et al., 2019). This is because the high-throughput emulsion techniques are typically aimed at characterising larger numbers of cells, which generally precludes deep sequencing, which would be expensive at that scale.

As a result, the workflows and processes from bulk RNA-seq cannot be directly utilised for the analysis of single-cell RNA-seq data. Instead, novel computational methods are required, specifically for processes that are directly affected by the high dropout rates or increased heterogeneity of single-cell data (Bacher and Kendziorski, 2016).

Emulsion-based single-cell RNA-seq data analysis is a multi-step process (Poirion et al., 2016). While numerous tools are available for every step of each process, there are also software packages that implement many, if not all, steps of the analysis workflow, such as Seurat (Satija et al., 2015) and scanpy (Wolf et al., 2018). In the first step of data analysis, the data is pre-processed and aligned to genomes, to convert the raw data generated by sequencing machines into matrices of read counts. The data is then subject to a quality control (QC) step to exclude poor-quality data from single cells. Here, three important factors are typically examined, namely the number of counts per barcode (count depth), the number of genes per barcode and the fraction of counts from mitochondrial genes per barcode (Luecken and Theis, 2019). Barcodes that are outliers can correspond to barcodes from dying cells, cells with poor mRNA recovery or low efficiency of cDNA production, or to doublets, where two cells were encapsulated in a single droplet (Haque et al., 2017), and these should be excluded.

The data is then normalised to account for variation in sampling between cells.

This variation can be caused by variability in the capture, reverse transcription and sequencing of mRNA, such that identical cells may have varied count depths. The count data does not account for this, but normalisation addresses this by, for example, scaling the count data to gather relative gene expression abundances within different cells (Luecken and Theis, 2019). This step has been found to have a large impact on single-cell RNA-seq analyses (Vieth et al., 2019).

Next, one can also account for and regress out the biological confounding factors, such as cell cycle, apoptosis (Stegle et al., 2015) and technical batch effects (Butler et al., 2018) that were previously mentioned.

Subsequently, the data can be analysed in a variety of ways, depending on the biological question of interest. The data can be clustered to identify different cell types or populations of interest, developmental trajectories can be mapped or differentially-expressed genes between pre-defined groups can be identified.

There exist various methods for the identification of differentially-expressed genes, with both methods developed for bulk RNA-seq and single-cell RNA-seq having been utilised for the analysis of single-cell RNA-seq data. The single-cell RNA-seq analyses packages, such as MAST and single-cell differential expression (SCDE), have been developed specifically to account for frequent dropout events and claim to be more sensitive in detecting differentially-expressed genes than the bulk RNA-seq methods (Finak et al., 2015; Kharchenko et al., 2014). However, some analyses have rather observed that methods designed specifically for single-cell RNA-seq data do not necessarily perform better than those for bulk RNA-seq data (Soneson and Robinson, 2018; Vieth et al., 2019; Wang et al., 2019).

Another challenge of analysing single-cell RNA-seq data comes from the sheer multi-dimensionality of the data generated. One has to examine the tens of thousands of genes, each of which can be considered as a single dimension, and consider their expression in hundreds to thousands of cells (Haque et al., 2017). Many approaches combat this by reducing the 'multi-dimensionality' of the data into smaller numbers of dimensions that can be easily visualised and analysed. For example, Principal Component Analysis (PCA) is a linear transformation method that reduces the dimensionality of data, typically from over 20,000 genes to 10 - 100 different principal components (Haque et al., 2017; Kiselev et al., 2019). PCA works to find the axis that explains the maximum amount of variance (the first principal component), before finding orthogonal axes that capture the greatest amount of remaining variance, to define the second and third principal component, and so on. This makes it fast, deterministic and linear, but means that it cannot capture non-linear relationships. In addition, PCA tends to be disproportionately affected by outliers (Amezquita et al., 2020).

In contrast, there are other more flexible non-linear techniques, such as t-distributed stochastic neighbor embedding (t-SNE) (Maaten and Hinton, 2008) and

Uniform Manifold Approximation and Projection (UMAP) (McInnes et al., 2018), which are capable of capturing non-linear dependencies. These work by trying to find a low-dimensional representation that continues to preserve relationships between neighbours in high-dimensional space (Amezquita et al., 2020). However, this comes at the expense of increased computational effort, and the need for the user to define more parameters, for which guidelines are not necessarily well-defined (Kiselev et al., 2019). It has been generally suggested that a range of different values be tested for the parameters, so as to not draw inaccurate conclusions (Amezquita et al., 2020). When comparing t-SNE and UMAP, UMAP aims to preserve more global structure, but at the expense of resolution within each cluster, while t-SNE is particularly suited for high-dimensional data. In addition, UMAP is computationally faster than t-SNE, making it increasing popular today (Amezquita et al., 2020).

The dimensionality reduction methods mentioned above can be used for both visualisation and summarisation (Luecken and Theis, 2019). Visualisation refers to optimally representing the dataset in two or three dimensions, which is useful for further understanding data, as well as for the preliminary detection of batch effects, in the case of PCA. In contrast, summarisation refers to using dimensionality reduction to aid downstream analysis. This both reduces noise and reduces the downstream computational work (Amezquita et al., 2020). In addition, this may reveal biologically interesting information. For example, the analysis of principal components by Macosko *et al.* revealed that the top principal components of a particular dataset were dominated by genes involved in the cell cycle, enabling thorough characterisation of cell-cycle gene expression (Macosko et al., 2015). The different techniques outlined above may thus be more suited for different applications, for example, Luecken and Theis recommend UMAP for exploratory visualization and PCA for general purpose summarization (Luecken and Theis, 2019).

Cells can then be organised into clusters, which permits data exploration and the deduction of member cell identity (Luecken and Theis, 2019). This takes place by grouping cells based on similarity of gene expression profiles. There are a variety of algorithms used for clustering, with one popular method being k-means clustering, which divides cells into k clusters in an iterative process (MacQueen, 1967). This approach requires one to input k, the expected number of clusters, which may not be known beforehand. In addition, k-means clustering is biased towards similarly-sized clusters, which may result in rare cell types being hidden within larger groups (Kiselev et al., 2019). In contrast, many single-cell analysis platforms currently use the Louvain algorithm on single-cell k-Nearest Neighbour (KNN) graphs (Blondel et al., 2008), which is capable of detecting communities as groups of cells that have more links between them than expected from the overall number of links that cells have. Clusters may then be manually annotated or annotated with the aid of reference databases such as the Mouse Brain Atlas (Zeisel et al., 2018) or

the Human Cell Atlas (Regev et al., 2017). This latter approach can be carried out by utilising either database-derived marker genes, or by using full gene-expression profiles from the relevant databases (Luecken and Theis, 2019).

While the above methods of clustering, dimensionality reduction and analysing differential expression frequently do consider and model the sparsity of single-cell RNA-seq data, the problem of sparsity can also be dealt with by "imputing" values for the observed zeros that better reflect the true gene expression levels (Lähnemann et al., 2020).

Various types of imputation have been used to infer gene expression levels. Some imputation methods adjust gene expression levels based on expression in "similar" cells, aiming to reduce the noise present (Dijk et al., 2018). Other methods use probability to determine which observed zeros are biological or technical, and remove technical variation while preserving biological variation (Huang et al., 2018; Li and Li, 2018). This is carried out by using the expression of other genes, or of the same genes in other cells as predictors. A separate class of methods utilises the generation of low-dimensionality representations to impute the data, as these methods enable the reconstruction of the observed data matrix from simplified representations (Pierson and Yau, 2015; Risso et al., 2018).

Imputation can be problematic when the imputations are entirely reliant on information present within the imputed dataset, as this can lead to circularity that artificially amplifies signals present within the dataset, leading to inflated correlations and false positives (Andrews and Hemberg, 2019). Therefore, any conclusions derived from imputed data should be subjected to thorough experimental validation (Luecken and Theis, 2019). This problem can also be overcome by reference to cell atlases, many of which are in the process of being developed (Almanzar et al., 2020; Regev et al., 2017), or by concurrently exploring complementary data types (Lähnemann et al., 2020).

Evidently, there are a plethora of assorted tools for the different steps of single-cell RNA-seq data analysis, with various tools and pipelines differentially suited for diverse applications. As this is an active research area, one expects the introduction of new tools in the future. In particular, it is thought that tools utilising deep learning workflows and tools for single-cell omic integrations may be exceptionally disruptive to current analysis pipelines (Luecken and Theis, 2019).

1.6.6 Targeted sequencing within emulsion-based single-cell transcriptomic analyses

Single-cell sequencing experiments can be costly, with costs increasing with both the number of cells sequenced and with sequencing depth, specifically the number of reads per cell. Thus, there have been multiple analyses on how best to optimise experimental design in the face of the trade-off between sequencing depth and cell

number in single-cell transcriptomic studies (Svensson et al., 2019; Zhang et al., 2020).

In many transcriptomic analyses, fewer than 1000 genes are variable and are responsible for the clustering of expression profiles (Birey et al., 2017). In addition, it has been shown that around 2% of genes consume over 50% of sequencing reads, while many genes with important biological functions are frequently expressed at lower levels (Replogle et al., 2020). It has also been demonstrated that many transcriptional states can be decently inferred from the gene expression of a subset of genes (Subramanian et al., 2017). This suggests that a targeted approach focusing on a set of particular genes could improve throughput and sensitivity while keeping sequencing cost-effective (Schraivogel et al., 2020).

This can be easily achieved in InDrop, which permits the straightforward manufacture of beads with customised capture sequences (Zilionis et al., 2017). This is possible as primer synthesis on the InDrop hydrogel beads are carried out in a stepwise manner, with two barcode segments, the UMIs and poly-dT sequences being added via sequential split-and-pool synthesis steps (Fig. 1.7A). By replacing the poly-dT sequences with a pool of gene-specific sequences, and by adding these via an additional synthesis step, transcriptomic sequencing can be carried out in a targeted manner (Zilionis et al., 2017).

An alternative technology, droplet-assisted RNA targeting by single-cell sequencing (DART-seq), has been developed to enable the concurrent analysis of the transcriptome and specific RNA amplicons, and has been demonstrated to work with Drop-seq beads (Fig. 1.7B). This is made possible via the enzymatic attachment of custom primers to a subset of poly-dT sequences on Drop-seq beads, enabling the specific capture of certain mRNAs. The DART-seq system has been utilised for characterisation of non-A-tailed transcripts of an RNA virus together with whole transcriptome analysis of the infected cells, and for the characterisation of paired variable heavy and light chain antibody sequences alongside whole transcriptomes of B-lymphocytes (Saikia et al., 2019).

Another possibility for targeted single-cell transcriptomic sequencing employs gene-specific primers to selectively amplify genes of interest following cDNA generation (Fig. 1.7C). This is exemplified by Constellation Drop-seq, which utilises a single primer per gene (Vallejo et al., 2019), and targeted Perturb-seq (TAP-seq), which utilises two different gene-specific primers per gene (Schraivogel et al., 2020). The former has been demonstrated with Drop-seq, while the latter has been utilised together with both Drop-seq and the 10X Genomics platforms. TAP-seq enabled the identification of 18 different cell subtypes within murine bone marrow with merely 100 sequencing reads per cell, which is an approximately 7- to 12-fold lower sequencing depth than what would be required to produce quantitatively similar results via whole transcriptome sequencing (Schraivogel et al., 2020).

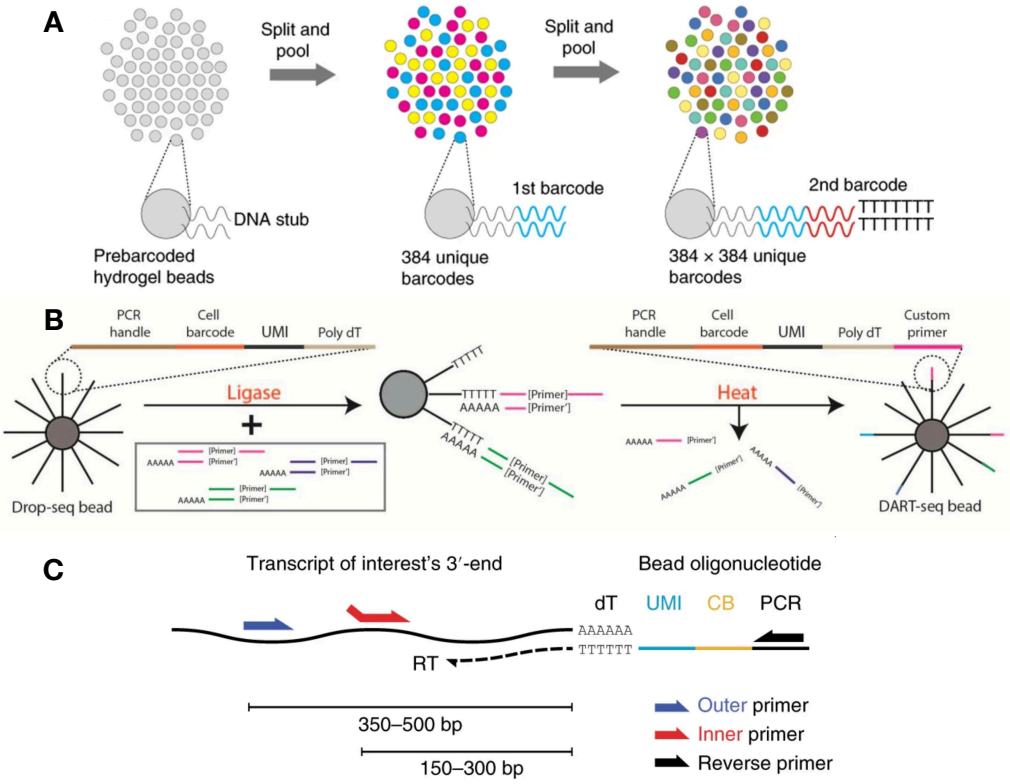


Figure 1.7: Targeted sequencing in emulsion-based single-cell transcriptomic analyses. **A.** Illustration of the split-and-pool approach used for on-bead oligonucleotide generation in InDrop. The approach is highly adjustable, where the addition of an additional synthesis step and the replacement of the poly-dT sequence with gene-specific sequences enables targeted sequencing. Figure reprinted from Zilionis et al., 2017 with permission from Springer Nature. **B.** Workflow for converting Drop-seq beads to DART-seq beads for targeted sequencing. Double-stranded probes containing custom primer sequences and poly-dA ssDNA overhangs anneal to the poly-dT segments of a subset of oligonucleotides present on Drop-seq beads. These probes are ligated to the oligonucleotide sequences present on the bead, and the complementary strands are removed via heating. The presence of both custom primers and oligo-dT tails permits simultaneous analysis of targets of interest together with whole transcriptome analysis. Figure reprinted from Saikia et al., 2019 with permission from Springer Nature. **C.** Illustration of the targeted Perturb-seq (TAP-seq) approach. Two gene-specific primers (outer and inner primers) are utilised to selectively amplify each gene of interest, after reverse transcription (RT) is carried out. The bead oligonucleotide illustrated is representative of both Drop-seq and 10X Genomics bead oligonucleotides, and includes common features of both, including unique molecular identifiers (UMI), cell barcodes (CB) and sequences involved in PCR amplification (PCR). Figure reprinted from Schraivogel et al., 2020 with permission from Springer Nature.

Alternatively, hybridisation-based target enrichment has been utilised to enrich for hundreds of selected transcripts. This has been demonstrated by direct-capture Perturb-seq, where hybridisation baits against 978 genes from the L1000 landmark genes (Subramanian et al., 2017) were designed and used for hybridisation capture (Replogle et al., 2020). This increased the percentage of mRNA molecules aligning to target genes from 6% without enrichment to 87% after target enrichment,

corresponding to an increase of over 14-fold. This suggests that sequencing at a 10-fold lower depth would still result in the enriched library containing more UMIs per cell for most targeted genes.

1.6.7 Emulsion-based single-cell cell surface proteomics

Following the establishment of emulsion-based single-cell transcriptomic techniques, various platforms for single-cell proteomic analyses capable of analysing membrane protein expression emerged (Peterson et al., 2017; Shahi et al., 2017; Stoeckius et al., 2017b).

Many of these technologies work in a similar manner - antibodies specific for the proteins of interest and labelled with oligonucleotide tags are used to stain cells, prior to encapsulation, cell lysis, barcoding of antibody tags (and cDNA) and library construction. Most methods, specifically CITE-seq (cellular indexing of transcriptomes and epitopes by sequencing) (Stoeckius et al., 2017b), REAP-seq (RNA expression and protein sequencing assay) (Peterson et al., 2017) and Feature Barcoding from 10X Genomics, are compatible with single-cell transcriptomic technologies such as Drop-seq or the 10X Genomics Single Cell 3' Gene Expression technology. This is made possible by the inclusion of a poly-dA tail at the 3' end of the conjugated oligo, which will bind to the poly-dT tails of the barcoding bead oligonucleotides, as shown below for the CITE-seq workflow (Fig. 1.8) (Stoeckius et al., 2017b). Alternatively, the introduction of complementary capture sequences on the oligonucleotide sequences of the antibodies and on the Gel Beads in 10X Genomics Feature Barcoding achieves the same purpose while enabling independent barcoding of the transcriptome and the proteome (Matuła et al., 2020).

The CITE-seq technology has been adapted by its creators to enable sample multiplexing that remains compatible with emulsion-based single-cell transcriptomics analyses (Stoeckius et al., 2018). Here, antibodies against ubiquitously expressed surface proteins are used. The same pool of antibodies, but with different oligonucleotide tags (known as cell hashtags), are used to stain different samples, before the cells from different samples are pooled and processed with the single-cell transcriptomic technology of choice. Sequencing both the cell hashtags and the transcriptomes ensures that the sample identity of each cell can be elucidated. As all samples are processed within one encapsulation experiment, this reduces batch effects. In addition, cell hashing allows one to identify most doublet events and eliminate them in the analysis. This facilitates the superloading of cells, where cells are loaded at a higher concentration than would normally be considered, due to the limitations of Poisson distribution. This permits the analysis of more cells for a given amount of reagents, reducing experimental costs.

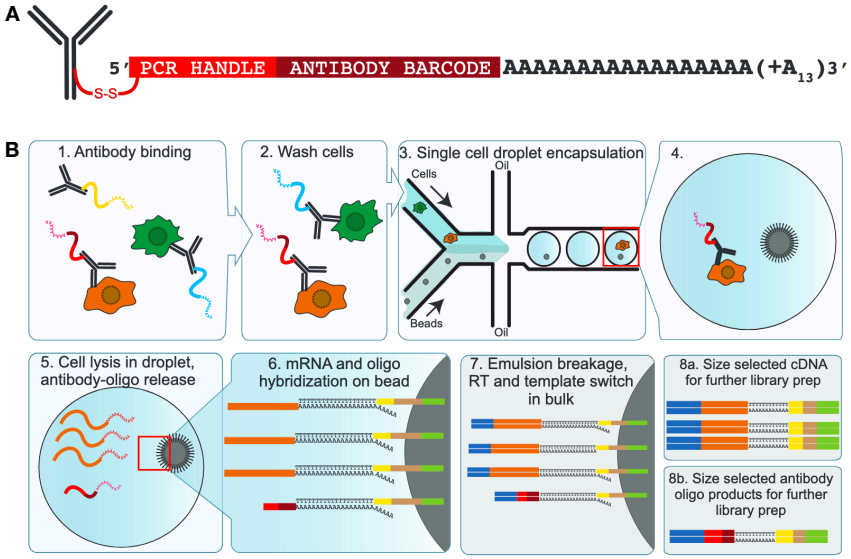


Figure 1.8: Antibody diagram and workflow used in CITE-seq and cell hashing. **A.** Illustration of the oligo-tagged antibodies used. Antibodies are bound to an oligonucleotide that contains a PCR barcode, an antibody barcode and a poly-dA tail. **B.** CITE-seq and cell hashing workflow. Cells are stained with antibodies and washed prior to mixing and encapsulation. Upon cell lysis, both the antibody-oligos and mRNAs are captured on beads. Subsequent processing takes place as appropriate for the respective emulsion-based single-cell transcriptomic technology – outlined here for Drop-seq. Figure reprinted from Stoeckius et al., 2017a with the permission of the authors.

1.6.8 Simultaneous profiling of genetic perturbations and transcriptomes in single cells

Single-cell transcriptomics has been utilised in screens involving genetic perturbations, namely where the perturbation conditions are encoded by genetic markers within the same cell in which the perturbation takes place. A subset of these screens, specifically Perturb-seq (Dixit et al., 2016), CROP-seq (CRISPR droplet sequencing) (Datlinger et al., 2017) and Mosiac-Seq (mosaic single-cell analysis by indexed CRISPR sequencing) (Xie et al., 2017) employ emulsion-based single-cell RNA-seq as a readout for CRISPR-based pooled genetic screens. The specific single guide RNA (sgRNA) associated with the perturbation is identified either by capturing the sgRNA itself within a Pol II transcript (CROP-seq) or by Pol II transcribed barcodes (Perturb-seq, Mosiac-seq), each of which is associated with a specific sgRNA sequence.

While the techniques studied various biological processes, ranging from LPS stimulation in mouse primary immune cells (Dixit et al., 2016) to enhancer function in human cell lines (Xie et al., 2017), most techniques nonetheless analyse an average of 80 - 150 transcriptomes per sgRNA and are capable of detecting both gene level regulation and signature and state-level regulation (Datlinger et al., 2017; Dixit et al., 2016; Hill et al., 2018; Xie et al., 2017).

As single-cell RNA-seq data is intrinsically noisy due to the high drop-out rate, all three screens carried out downsampling or saturation analyses, to assess the sensitivity and/or specificity of their readout as compared to bulk RNA-seq data or to a more comprehensive dataset, specifically for different numbers of transcriptomes per perturbation. This naturally remains dependent on effect size of the perturbation. In Mosaic-seq, they have mapped the percentage recovery of differentially expressed genes as a function of the number of single cells analysed. With 100 cells, they were able to pick up approximately 35.7% of differentially expressed genes seen in bulk RNA-seq (Xie et al., 2017). Pilot experiments in Perturb-seq found that with 100 single cells per guide, the sensitivity and specificity of detecting the correct genes regulated by the perturbation (as seen in bulk RNA-seq) were 80% and 90% respectively, but that signature and state-level regulation should be detectable with a mere tens of cells per perturbation, together with 400 transcripts per cell (Dixit et al., 2016). In CROP-seq, they estimated that 15% (850/5,798 transcriptomes) of the dataset would be sufficient to obtain concordant gene expression signatures seen with their current data (Datlinger et al., 2017), which is approximately 12 - 13 cells per perturbation.

These illustrate that for different biological processes, single-cell transcriptomics is capable of detecting differential gene expression induced by a specific perturbation that can be concurrently detected via sequencing. The actual number of cells required to pick up a specific effect would naturally differ with the strength of the perturbation, the genes examined and the sequencing depth (Datlinger et al., 2017). This raises the question of how many single cells would be required to detect a given perturbation, and if a perturbation could be identified from its effect on the transcriptome of a single cell. This thesis aims to investigate this in the context of drug, antibody and cell-based perturbations.

1.7 Cell-cell interactions

Cell-cell interactions are essential in many cellular processes, including proliferation, differentiation, survival, apoptosis, migration and the coordination of responses to stimulation (Guo et al., 2013). The misregulation of cell-cell interactions have been associated with disease, including neurodegenerative diseases (Beer et al., 2008) and cancer (Jain et al., 1996). Cell-cell interactions can take place between cells of the same cell type (homotypic) or different cell types (heterotypic), and can occur either via direct cell contact or by the transmission of a signal, such as via soluble factors, electrical signals or mechanical cues (Zervantonakis et al., 2011).

The study of cell-cell interactions has largely taken place *in vitro*, as such experiments are experimentally less complex and enable standardisation and simplification of experimental conditions to eliminate potential confounding factors found in experiments conducted *in vivo* (Guo et al., 2013). Cell-cell interactions have

been studied in bulk with Transwell plates, where two different cell populations are separated by a porous filter, enabling the study of paracrine interactions but not those involving direct contact. In addition, other co-culture assays have also been used, using Petri dishes, solid supports and bioreactors (Goers et al., 2014). However, these methods are limited by the lack of control over the loading of cells and the concurrent lack of spatiotemporal regulation of the microenvironment that cells are exposed to (Guo et al., 2013).

In contrast, microfluidics affords experimentalists with higher control over the microenviroment that cells are exposed to (Velve-Casquillas et al., 2010) and permits spatiotemporal regulation of biochemical cues (Vu et al., 2017). In addition, such systems permit better manipulation of cell seeding (Matsue et al., 1997) and modulation of cell patterning (Park and Shuler, 2003). The small volumes used also allows the isolation and further study of soluble factors utilised in communication, as these can be found at functional concentrations (Guo et al., 2013).

Microfluidics permits both the study of population-level cell-cell interaction studies, which utilise larger chambers to culture larger numbers of cells, as well as single-cell level cell-cell interaction studies, which mostly pair and study two cells (Dura et al., 2015; Dura et al., 2016). Both types of studies will be discussed below.

1.7.1 Population level interaction studies with microfluidics

The versatility afforded by microfluidic systems allows the design of different chambers or channels in which different cell types can be seeded. These can be made permeable only to soluble factors or separated by pneumatic valves to specifically study how cells communicate indirectly (Xu et al., 2010), or they can be separated by mere pillars to enable cells to be in direct contact. Such systems have been used to investigate a plethora of different biological systems, including hepatocyte-endothelial cell interactions (Sudo et al., 2009) and neuromuscular junction formation (Zahavi et al., 2015).

These systems benefit from recent advances in three-dimensional (3D) cell culture that better mimic the extracellular matrix *in vivo*. For example, Bersini *et al.* has developed a microfluidic 3D *in vitro* model of a vascularised bone-like microenvironment, by using a collagen gel embedded with human mesenchymal stem cells and lined with endothelial cells. This was then incubated with invasive breast cancer cells, to study the process by which breast cancer metastasises into bone (Bersini et al., 2014).

These systems are frequently analysed via real-time imaging and bulk analyses of secreted soluble factors, which enable the investigation of cell-cell interactions at the population level in a more quantitative and regulated manner than before.

1.7.2 Single-cell interaction studies with microfluidics

While the study of cell-cell interactions at a population level will reveal much about the interactions, the large numbers of cells would mask any heterogeneity in cell responses. In contrast, the study of two single cells that are paired in a regulated manner enables the study of such heterogeneity with a high level of detail. In addition, it is of interest to screen various libraries, such as libraries of antibodies, T-cell receptors or engineered factors, for their effect on cell-cell interactions, and the utilisation of pairs of single cells makes this feasible at a high throughput.

Pairs of single cells can be brought together by the use of nanowells. This has been used by Yamanaka *et al.* to co-incubate natural killer (NK) cells together with K-562 target cells. The cells were imaged at various time points to assess if cell killing took place and cytokine and chemokine secretion was measured by the utilisation of a capture antibody-coated glass slide that was incubated in contact with the nanowell array. This produced a matched microarray of secreted proteins that were then detected by the use of fluorescently labeled antibodies and imaging, and correlated back to specific nanowells. The authors found that the secretion of IFN-γ from NK cells is associated with its motility when it contacts the target cell, but not with the outcome of cell killing (Yamanaka et al., 2012).

Cells can also be brought together by co-encapsulation. Sarkar *et al.* studied the interactions between T-lymphocytes and dendritic cells (DCs) by imaging to assess cell interaction and by measuring calcium levels via fluorescence microscopy. They noted an increase in T-lymphocyte calcium levels in both directly interacting DC-T-lymphocyte pairs and those in the same droplet but that did not appear to make contact, suggesting the presence of contact-dependent and independent processes in T-lymphocyte activation (Sarkar et al., 2015).

Microfluidic hydrodynamic trapping can be utilised to trap and pair cells in a deterministic manner at a 1:1 cell ratio (Dura et al., 2015). Here, cells are sequentially loaded into traps, which hold the cell pairs in place even if the surrounding media is changed. Dura *et al.* examined the interaction of $CD8^+$ T-lymphocytes with antigen-loaded B-lymphocytes and the subsequent effects of ionomycin stimulation. By examining cytosolic calcium mobilisation and the expression of CD8 (on T-lymphocytes) and pMHCI and MHCII-eGFP (on B-lymphocytes), they found that there is a heterogeneity in the response of T-lymphocytes, where some cells respond to both stimuli (double responders), some respond to only ionomycin activation (single responders) while some did not respond entirely (Dura et al., 2015). A subsequent analysis from the same group tracked more parameters, with both on- and off-chip culturing, imaging, staining and single-cell sequencing shown to be possible (Dura et al., 2016).

Such microfluidic single-cell interaction assays enable the detection of cell heterogeneity in cell-cell interactions. However, most of these assays are low

throughput, due to the dependence on imaging and single-cell manipulations. Additionally, many assays are only able to monitor a few parameters at once, which may not provide enough information for a complete understanding of the complexity observed in many cell-cell interaction processes.

While there exist many techniques to study cell-cell interactions at the single-cell level, most are only able to track a few different parameters at a time and in a low-throughput manner. At the same time, there exist various emulsion-based methods for high-throughput transcriptomic analyses of single cells, where the expression of all genes within the cells can be assessed, but these have not been applied to study cell-cell interactions. Therefore, the development of a method that permits high-throughput transcriptomic analyses of cell-cell interactions is of great interest and is one goal of this thesis.

1.8 Monoclonal antibodies

One particularly interesting category of cell-cell interactions involves soluble antibodies secreted by antibody-secreting cells, together with target cells that express surface proteins that these antibodies bind to. These heterotypic interactions are particularly important given the increasing important role that monoclonal antibodies (mAbs) play in modern medicine, with seven of the top ten best-selling drugs in 2019 being mAbs (Urquhart, 2020a), and with mAb sales expected to reach nearly USD 200 billion by 2024 (Dealmakers, 2019). Additionally, 12 new mAbs were approved by the United States Food and Drug Administration (FDA) in 2018, representing 20% of the total approved drugs in that year, and half of these are expected to be blockbusters and to generate peak sales of over USD 1 billion by 2024 (Mullard, 2019). These illustrate the continued importance of mAbs in modern medicine.

While this chapter will only focus on monoclonal antibodies, it must be noted that there are also other promising antibody-derived therapeutics in development or on the market, including bispecific antibodies, which are antibodies that have been designed to target two different epitopes (Labrijn et al., 2019), and immunoadhesins, which are chimeric molecules combining the functional domain of a binding protein with immunoglobulin constant domains (Ashkenazi and Chamow, 1995).

1.8.1 Generation, structure and function of antibodies

B-lymphocytes are a crucial part of the adaptive immune system, with their function carried out by the production of clonally diverse immunoglobulin (Ig) receptors that are capable of precisely recognising antigenic epitopes with nano- to picomolar affinities (LeBien and Tedder, 2008). These immunoglobulins (Ig) can be present as surface immunoglobulins on the surface of B-lymphocytes, or secreted as soluble antibodies (Hoffman et al., 2016).

The diversity of antibodies and B-cell receptors (BCRs) expressed within an individual immune system is immense, and this is made possible by V(D)J recombination, a somatic recombination process in B-cell development. The process brings together one variable (V), one diversity (D) and one joining (J) segment of the *IGH* locus to form an exon in the heavy chain Ig gene, with the same taking place for the V and J segments of the light chain. Additional diversity can be introduced by random insertions or deletions at the segment junctions (Hoehn et al., 2016). In addition, BCRs improve antigen binding during infection via affinity maturation, a combination of both somatic hypermutation (Li et al., 2004) and clonal selection, such that the remaining B-cell progeny secrete antibodies with higher affinities for the antigen (Peled et al., 2008).

While all antibody molecules made by an individual B-lymphocyte typically have the same antigen-binding site, the sheer number of different B-lymphocytes with diverse antibodies and BCRs generated by the processes outlined above, explains the polyclonality of antibodies obtained from an organism. It has been estimated that the potential diversity of the naïve repertoire is over 10^{18} (Elhanati et al., 2015), yet this does not consider somatic hypermutation, and thus the true potential diversity is potentially much larger. In contrast, monoclonal antibodies utilised in medicine are produced by a single clone of B-lymphocytes, which typically produce a single species of antibody against a single epitope. This means that such monoclonal antibodies are homogeneous, monospecific and characterisable for therapeutic purposes (Castelli et al., 2019).

Five different types of immunoglobulins are produced by B-lymphocytes, namely IgG, IgA, IgM, IgD and IgE. These differ in the type of heavy chain present in the molecule and these differences permit the immunoglobulins to function in different types and different stages of immune responses (Schroeder and Cavacini, 2010). For example, IgM is excellent at activating complement, while IgE binds Fc receptors to activate mast cells (Hoffman et al., 2016). The different types of immunoglobulin also differ in their affinities, where high molecular weight immunoglobulins (IgM) have lower affinities for the antigens, as compared to IgG and IgA, which generally have affinities that are orders of magnitude higher, due to affinity maturation (Eisen, 2014). All therapeutic antibodies that are currently used in the clinic are immunoglobulin G (IgG) monoclonal antibodies, of which four different isotypes exist (IgG1, IgG2, IgG3, and IgG4) (Ryman and Meibohm, 2017). IgG1 is the most abundant isotype in serum, constituting 60% - 75% of serum IgG (Hoffman et al., 2016).

A single immunoglobulin G monoclonal antibody consists of four protein chains, two identical heavy chains and two identical light chains, which are linked via disulfide bonds (Fig. 1.9A). The N-terminal regions of the chains form the variable (V) region, which includes the antigen-binding complementarity-determining regions (CDRs), while the C-terminal regions form

the constant (C) regions (Awwad and Angkawinitwong, 2018). The IgG can also be divided into the Fab (fragment antigen-binding) and Fc (fragment crystallisable region) regions, with the Fab regions permitting the identification of different antigens by different antibodies, and the Fc regions regulating various effector functions (Porter, 1958).

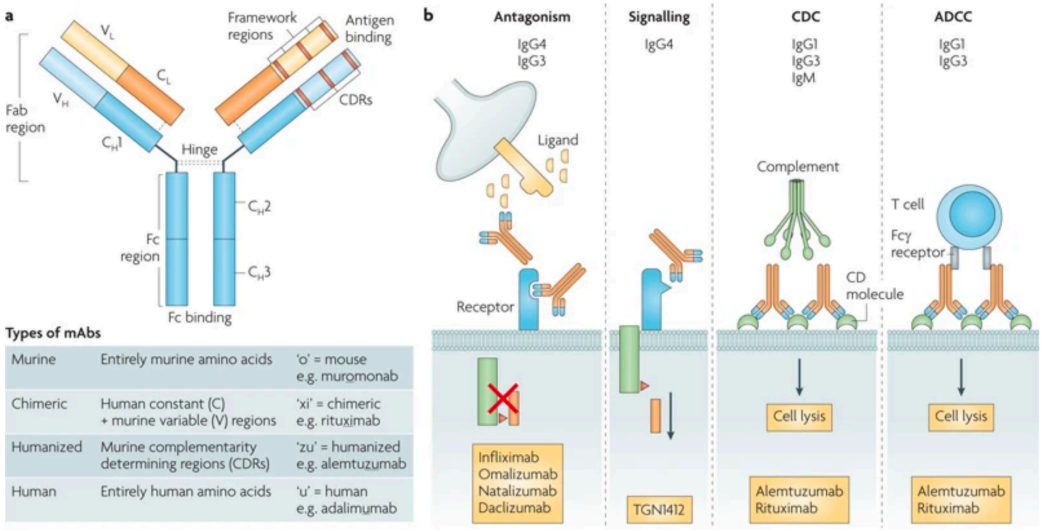


Figure 1.9: Monoclonal antibody structure and function. **A.** Schematic of an immunoglobulin G (IgG) monoclonal antibody (mAb). The constant heavy (C_H) and constant light regions (C_L), and the variable heavy (V_H) and variable light regions (V_L) are indicated. The table shows the various types of mAbs present, including murine, chimeric (human constant (C) with murine variable (V) regions), humanised (with only murine complementarity-determining regions (CDRs)) and human mAbs. **B.** Mechanisms of action of mAbs. The mechanisms of actions of antibodies are provided together with examples of therapeutic mAbs that act via those mechanisms (orange box). Antagonism and signalling are controlled by specific CDRs within the Fab region, where antibodies can specifically bind to a target (either ligand or receptor) and block (Antagonism) or induce signalling (Signalling). In contrast, other functions of mAbs are controlled by the Fc region, namely complement-dependent cytotoxicity (CDC) and antibody-dependent cell-mediated cytotoxicity (ADCC). mAbs can cause cell lysis of T- or B-lymphocytes via complement activation (CDC) or bind to Fc receptors and trigger cell lysis (ADCC). Figure reprinted from Hansel et al., 2010 with the permission from Springer Nature.

Different types of monoclonal antibodies exist with different portions of antibodies being derived from murine or human proteins (Fig. 1.9A). Early mAbs, such as Muromonab-CD3, the first mAb approved for use in humans, were fully murine, but these posed immunogenicity problems, where patients developed human anti-murine antibodies, resulting in rapid clearance of the drug and in anaphylactic shock in some patients (Abramowicz et al., 1992; Jaffers et al., 1986). In order to minimise immunogenic problems, the human content of therapeutic antibodies has been increased, with the development of chimeric (mouse/human) and humanised antibodies (human antibodies containing mouse CDRs) with the

aid of genetic engineering. Here, the mouse antibody variable regions or the mouse CDRs are transferred to human IgGs respectively, reducing the non-human content of antibodies to 30% and 5 - 10% respectively (Yamashita et al., 2007), in order to minimise the immunogenicity of such mAbs. In addition, technologies have been developed that permit the discovery of fully human mAbs, namely phage display and the immunisation of transgenic mouse models that express human IgG isotypes (Yamashita et al., 2007). More information about these methods can be found in Section 1.9.

mAbs can act via various mechanisms (Fig. 1.9B). As previously mentioned, the Fab domain enables recognition of different antigens, which permits it to block ligand-receptor interactions (Antagonism) or to induce receptor signalling in the absence of ligands (Signalling) (Hansel et al., 2010). Additionally, the Fc region of mAbs permits binding to Fc receptors (FcR), which initiates complement-dependent cytotoxicity (CDC) or antibody-dependent cellular cytotoxicity (ADCC). These vary with Ig subclass, where the isotypes IgG1 and IgG3 are the most potent at activating complement, and IgG1 is highly effective at promoting ADCC, such that IgG1 is widely used to trigger cell death, such as in oncology-related applications (Weiner et al., 2010). In contrast, due to its reduced Fc effector function, IgG4 is commonly used if cytotoxicity is not desired, and further reduction of cytotoxicity can be induced via modification of the Fc region by genetic engineering.

1.8.2 Pharmacology of monoclonal antibodies compared to small molecule drugs

mAbs differ from small molecule drugs in various aspects. The high specificity of mAbs permits precise targeting and minimises off-target effects (Leader et al., 2008). While small molecule drugs have struggled to modulate protein-protein interactions, mAbs perform better in modulating such interactions (Mabonga and Kappo, 2019). In addition, the neonatal Fc receptor (FcRn) reduces the lysosomal degradation of IgG antibodies, including therapeutic mAbs (Kim et al., 2007), resulting in IgG1, IgG2 and IgG4 having a half-life of around 18 - 21 days, which is significantly longer than the typical half-life of other therapeutic molecules (Kontermann, 2011), thus lower dosing frequencies are sufficient (Ovacik and Lin, 2018).

mAbs are approximately 150 kDa in size, making them comparatively larger than small molecule drugs, which are typically between 200 - 500 Da in size (Wan, 2016). This generally limits them to extracellular targets, given their limited membrane permeability (Trenevska et al., 2017), although various strategies to intracellularly target mAbs are being explored (Slastnikova et al., 2018). In addition, the large size, limited membrane permeability and gastric degradation of mAbs restricts oral administration, such that they are typically administered via intravenous (IV), subcutaneous (SC), or intramuscular (IM) injections (Wang et al., 2008). Their large size also results in a slower distribution into tissue, with maximal

plasma concentrations noted only approximately 1 – 8 days following subcutaneous (SC) or intramuscular (IM) injection (Kamath, 2016).

1.8.3 Monoclonal antibodies in the clinic

mAbs in the clinic have been largely used for the treatment of cancer and autoimmune diseases, and to a smaller extent, against infectious diseases. In the treatment of cancer, mAbs targeting tumour antigens, such as those overexpressed in cancer cells and involved in unregulated cell division, have been successful in reducing tumor growth rate, inducing apoptosis and sensitising tumours to chemotherapy (Scott et al., 2012). For example, trastuzumab and pertuzumab bind HER2 (ERBB2), a receptor overexpressed in 30% of invasive breast cancers and in some adenocarcinomas of the lung, ovary, prostate and gastrointestinal tract (Chen et al., 2003). A combination therapy of trastuzumab and pertuzumab has been shown to be more effective in the treatment of HER2+ metastatic breast cancer than treatment with the single mAbs (Fabi et al., 2016). In addition, other modalities, such as bispecific antibodies (Brinkmann and Kontermann, 2017) , and antibodies conjugated to cytotoxic organic compounds (Antibody-Drug Conjugates (ADCs)) (Beck et al., 2017), radionuclides (Mattes, 2002) and protein toxins (Alewine et al., 2015) are increasingly utilised in the treatment of cancer (Almagro et al., 2017). The highly selective nature of mAbs results in fewer side effects compared to traditional chemotherapy (Pento, 2017), making them an attractive alternative for the treatment of certain cancers.

mAbs have also been utilised in the treatment of autoimmune diseases, where they are usually used to target various components of the immune system to reduce excessive disease-causing immune responses (Castelli et al., 2019). For example, mAbs targeting TNF-α, a cytokine important in autoimmunity, has been used in the treatment of various immune diseases including rheumatoid arthritis and Crohn's disease (Castelli et al., 2019). In addition, a few mAbs have been approved for prophylaxis and treatment of both bacterial and viral targets, such as bezlotoxumab against *Clostridium difficile* toxin B and palivizumab to prevent severe lower respiratory tract disease due to respiratory syncytial virus (RSV) in high risk children (Pelfrene et al., 2019). While not many mAbs have been approved for the treatment of bacterial and viral infections, it is thought that mAbs could be beneficial in treating emerging infections, especially in the absence of suitable alternatives (Pelfrene et al., 2019; Salazar et al., 2017), illustrating the continued importance of mAbs in the clinic.

1.8.4 Antibody sequencing

Given the diversity of B-cell receptor (BCR) repertoires, the sequencing of B-cell receptors is crucial to understanding B-lymphocyte clonal evolution in various contexts, such as during infection, vaccination and autoimmune diseases (Seah

et al., 2018). In addition, the sequencing of antibody-secreting cells from immunised animals and humans can be used to elucidate information on antibody binding. Equally important is the fact that antibody sequencing plays a vital supporting role in antibody screening (as will be elaborated on in Section 1.9), as the identification and characterisation of hits frequently requires the cloning and sequencing of antibody variable regions.

Traditionally, antibodies have been identified via the use of qPCR, or sequenced via a combination of single-cell PCR and Sanger sequencing. For example, a previous proof of concept droplet microfluidic antibody screen utilised qPCR to study hybridoma enrichment (Shembekar et al., 2018), while the BCR repertoire was traditionally characterised by the use of a combination of FACS to isolate single B-lymphocytes, single cell PCR and Sanger sequencing (Tiller et al., 2008). However, qPCR requires pre-knowledge of the antibody sequences, while the use of Sanger sequencing limits the analysis to tens or hundreds of BCR sequences, which is a miniscule fraction of the BCR repertoire, which has been estimated to be larger than than 1 x 10^{13} in humans (Georgiou et al., 2014). This suggests that high-throughput sequencing methods, such as those utilised in next generation sequencing (NGS), should be employed to expand the sampling of the antibody repertoire and to characterise a large number of hits during high-throughput antibody screening. This high-throughput DNA sequencing of immunoglobulin genes has been termed Ig-seq (Georgiou et al., 2014), but refers to a collection of different methods.

Antibody sequencing with high-throughput sequencing methods is not trivial, firstly due to the extraordinary variability in BCR sequences (Hoehn et al., 2016). The reverse transcription of mRNA and corresponding PCR amplification of cDNA ensures the amplification of expressed antibody sequences, but the variability means that a mix of primers has to be used and optimised to maximise recovery while minimising amplification bias (Tiller et al., 2008).

The second complication is that antibody heavy and light chains are encoded by different genes. This means that the pooling and sequencing of B-lymphocytes results in the inevitable loss of endogenous heavy and light chain pairings. The sheer number of heavy and light chain variants means that the restoration of these pairings cannot take place via trial-and-error.

This has been partially resolved by Reddy *et al.* for the pairing of high abundance clones, by the pairing of the most dominant heavy and light chains according to their relative frequencies. 78% of the identified antibodies generated in this manner were antigen-specific, but it must be noted that the diversity of the light chains was lower than that of heavy chains, such that multiple pairings had to be generated and validated by affinity measurements (Reddy et al., 2010).

Alternatively, the topologies of the V_H and V_L phylogenetic trees have been utilised to pair heavy and light chain sequences (Zhu et al., 2013). The authors

observed that V_H and V_L genes within matching branches of the respective phylogenetic trees could be paired to give antibodies with significantly lower autoreactivity compared to those those from unmatched branches, suggesting that the phylogenetic matching of heavy and light chains could approximate natural pairings (Zhu et al., 2013).

A more precise method that is highly compatible with NGS methods is the introduction of cell barcodes to label sequences from individual B-lymphocytes prior to pooling. The labelling of heavy and light chains from each cell with the same barcode ensures that the heavy chain sequences can be traced back to their associated light chain sequence following next-generation sequencing. This is analogous to molecular barcoding utilised in single-cell genetic analyses, as explained in Section 1.6.1.

This has been carried out in microtitre plates, where a two-dimensional barcoded primer matrix consisting of 240 x 192 tags, capable of barcoding over 46,000 individual B-lymphocytes, was developed. Of 1,152 single B-lymphocytes analysed, including peritoneal cavity B-cells, splenic marginal zone B-cells and mature naïve B-cells, 33% yielded full-length heavy and light chain genes (Busse et al., 2014). However, the processing of thousands of single-cell PCRs in microtitre plates is relatively expensive, labour-intensive and time-consuming.

Alternatives have been developed to take advantage of the small assay volumes and high throughputs possible with microfluidics. For example, a chip containing 1.7×10^5 nanowells was utilised for the deposition of single B-lymphocytes together with magnetic poly-dT beads. Cell lysis and mRNA capture on beads was carried out in individual nanowells, with magnetic beads subsequently pooled, washed and encapsulated into droplets together with reagents for RT-PCR and overlap extension PCR. Overlap extension PCR was utilised to link the associated heavy and light chain fragments, ensuring that endogenous pairings can be deciphered via subsequent sequencing. After RT-PCR and overlap extension PCR, the emulsion was broken and the fused V_H:V_L cDNA fragments were recovered and analysed via paired-end Illumina sequencing, enabling the single-cell analyses of over 5×10^4 lymphocytes (DeKosky et al., 2013).

The technology has also been adapted into a fully emulsion-based technology, where droplets were utilised instead of nanowells. Using a flow-focusing device, single B-lymphocytes were encapsulated together with lysis buffer and poly-dT beads, enabling the capture of mRNA on the beads. The beads were then recovered and utilised for emulsion overlap extension RT-PCR as outlined above. This enabled the sequencing of over 2×10^6 B-lymphocytes in one experiment, with a pairing precision of over 97% (DeKosky et al., 2015; McDaniel et al., 2016). Similarly, emulsion overlap extension RT-PCR has been employed to generate DNA amplicons that encode natively-paired single-chain variable fragments (scFvs), which were then

expressed as yeast display libraries and used for the screening of antibody binding. NGS was subsequently used to identify the clones in both pre- and post-sort scFv libraries (Adler et al., 2017a; Adler et al., 2017b).

1.8.5 Antibody sequencing coupled with emulsion-based single-cell transcriptomic analyses

Various single-cell antibody sequencing methods have been developed, either as part of, or in conjunction with the emulsion-based single-cell transcriptomic technologies mentioned in Section 1.6.4.

It is worth noting that unaltered conventional emulsion-based single-cell transcriptomic technologies which capture mRNA transcripts at their 3′ ends, such as Drop-seq, InDrop and the 10X Genomics Single Cell 3′ Gene Expression technology, cannot pick up the variable regions of antibody genes. Such technologies capture mRNA via their poly-A tails, with the cell barcodes being located at the 3′ end of the cDNA. However, the variable regions of the antibody genes are located at the 5′ end of their respective mRNAs (Fig. 1.10A). The utilisation of fragmentation and short-read Illumina sequencing results in a failure to sequence the variable regions of antibody and BCR transcripts using these technologies. Therefore, a variety of technological modifications have been developed to ensure that the variable regions of antibody transcripts and the cell barcodes from the same fragments can be sequenced, such that the native pairing of the heavy and light chains can be identified and simultaneously associated with whole-cell transcriptomic data.

The 5′ capture of mRNA has been adapted for antibody sequencing, as has been outlined by Goldstein *et al.* and utilised by the Single Cell Immune Profiling workflow from 10X Genomics. While the overall workflow is similar to that of the 3′ Gene Expression technology, with the utilisation of deformable gel beads to capture mRNA, the nucleotide sequences used here are organised differently, with the oligonucleotides containing the cell barcodes, UMIs and the segments for template switching located on the gel beads, and the poly-dT primers added as free-floating primers in the droplets (Fig. 1.10B). This results in the capture of mRNA on beads at their 5′ ends, with the cell barcodes and UMIs located at the 5′ end of the captured mRNA. As such, upon the generation of cDNA, the variable regions (VDJ) are located closer to the cell barcodes and UMIs, enabling the targeted amplification of B-cell receptor sequences when BCR-specific primers are used (Goldstein et al., 2019; *Single Cell Immune Profiling*). This enables the generation of sufficiently short fragments that can be sequenced on an Illumina platform, to yield heavy and light chain variable region sequences associated with the relevant cell barcodes (Goldstein et al., 2019). Such antibody sequencing can also be run in conjunction with the 10X Genomics Single Cell 5′ Gene Expression workflow, enabling the elucidation of whole-cell transcriptomic data together with the associated antibody sequences (*Single Cell Immune Profiling*).

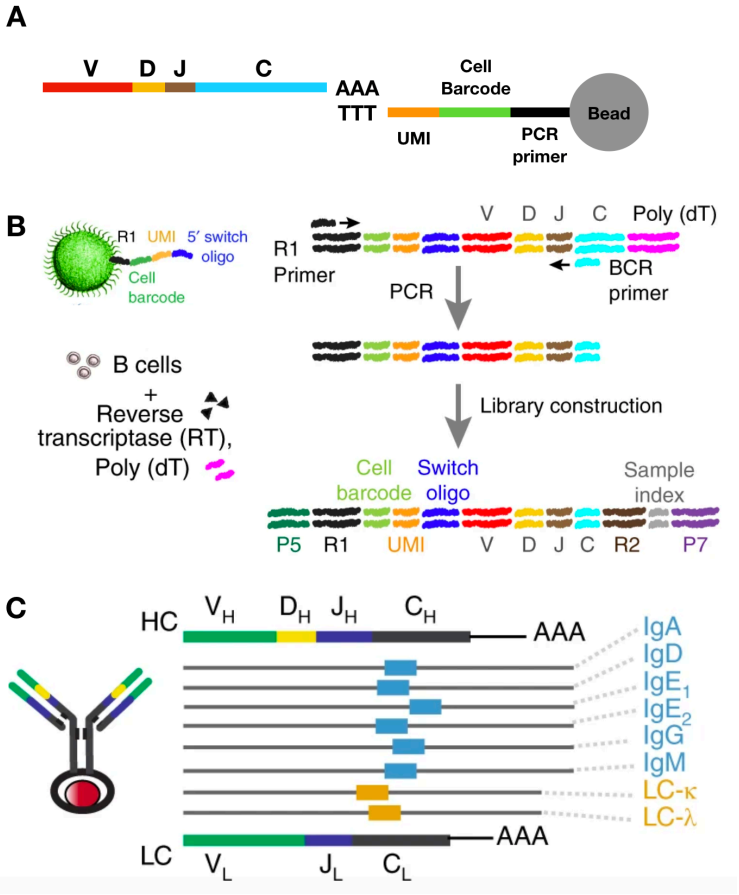


Figure 1.10: Problems with and strategies for antibody sequencing in parallel with single-cell transcriptomics. A. Capture of antibody-encoding mRNA via 3′ poly-A tails. The V, D and J regions of a transcript encoding an antibody heavy chain are indicated in red, yellow and brown respectively, with the constant region (C) shown in light blue. The mRNA is captured via hybridisation of the poly-A tail to the poly-dT sequence of bead-bound oligonucleotides. The UMI, cell barcode and PCR primer are shown in orange, green and black, respectively. Here, the constant region will separate the variable regions from the cell barcodes and UMIs. **B.** Schematic illustrating 5′ capture of mRNA and library construction for antibody sequencing. The beads utilised have oligonucleotides containing the R1 primer (black), cell barcode (green), UMI (orange) and the 5′ switch oligo (blue). These are co-encapsulated with B-cells (hollow brown circles), reverse transcriptase (black triangles) and poly-dT fragments (fuchsia). These result in fragments with the V, D and J segments (red, yellow and brown, respectively) located directly adjacent to the initial bead-bound segments. The targeted amplification with BCR primers and R1 primers ensure that only antibody cDNA is amplified and generates fragments for library preparation that have most of the constant region removed. Figure modified and reprinted from Goldstein et al., 2019 under the terms of the Creative Commons Attribution License (CC-BY 4.0). **C.** DART-seq primer design enables the specific sequencing of antibody heavy and light chains. The heavy (HC, top) and light chain genes (LC, bottom) are shown. Primers were designed to anneal to the 5′ ends of the constant regions, and are shown here for the heavy (blue) and the light chains (orange). These ensure the specific capture of antibody mRNA onto Drop-seq beads, with part of the constant region not being included in the cDNA generated. Figure reprinted from Saikia et al., 2019 with permission from Springer Nature.

DART-seq (droplet-assisted RNA targeting by single-cell sequencing), as previously explained in Section 1.6.6, permits targeted sequencing via the ligation of gene-specific oligonucleotides onto Drop-seq beads, enabling gene-specific capture of the mRNAs of interest (Saikia et al., 2019). By utilising oligonucleotides targeting the constant regions of human heavy and light isotypes (Fig. 1.10C), antibody mRNAs can be specifically captured, with a part of the constant region excluded from the cDNA generated, enabling the sequencing of both antibody sequences and cell barcodes with Illumina sequencing. DART-seq has been utilised to obtain the complete CDR3L+ and CDR3H regions from 21.3% (120/564) of immunoglobulin-transcript-positive primary B-lymphocytes (Saikia et al., 2019).

Another technology that enables simultaneous whole-cell and antibody gene sequencing is Repertoire and Gene Expression by Sequencing (RAGE-Seq), which has been demonstrated to be compatible with the 10X Genomics Single Cell 3′ Gene Expression technology. Here, full length cDNA sequences generated from the 3′ capture of antibody-encoding mRNAs are enriched via targeted hybridisation capture. These cDNA sequences are then subjected to long-read nanopore sequencing, enabling sequencing of both the 3′ cell barcode and the 5′ VDJ regions. Short-read Illumina sequencing is carried out in parallel for the remaining cDNA, and the data from both datasets can be linked via matching of the cell barcodes present within both datasets. This enables the accurate pairing of whole single-cell transcriptomic profiles with targeted full-length antibody mRNA sequences. The technology recovered paired immunoglobulin heavy and light chains for 42.6% (689/1619) of primary B-lymphocytes obtained from a lymph node (Singh et al., 2019).

The above three techniques utilise different workarounds to obtain the sequences of antibody variable regions and their associated barcodes, enabling the native pairing of heavy and light chain antibody sequences, together with association to the relevant single-cell whole-cell transcriptomic data. As T-cell receptors (TCRs) resemble antibodies and B-cell receptors in terms of protein structure and the genetic mechanism by which variability is produced (Charles A Janeway et al., 2001), one could envision easily adapting the above strategies for the sequencing of TCRs, and this has been done for the 10X Genomics Single Cell Immune Profiling workflow (*Single Cell Immune Profiling*) and RAGE-seq (Singh et al., 2019).

1.9 Antibody discovery

The crucial role played by monoclonal antibodies in medicine has necessitated the development of various technologies for antibody discovery. These technologies are capable of screening for antibody binding or antibody function, or for a combination of both. While current bulk antibody screening technologies have been successfully utilised for the discovery of therapeutic monoclonal antibodies, they continue to be

limited to only screening for binders (phage display) or are limited by time, cost and the diversity generated (hybridoma technology). Various microfluidic technologies, specifically those involving droplet microfluidics, have been developed in recent years to both act as viable alternatives for, and to complement traditional antibody discovery methods.

This section will outline non-microfluidic, microfluidic and droplet microfluidic antibody discovery methods, for screening for both antibody binding and function.

1.9.1 Non-microfluidic antibody discovery methods

Phage display is the most widely used technology for *in vitro* antibody selection, where it is capable of rapidly identifying antibody binders from diverse libraries (with repertoire sizes of up to 10^{10} to 10^{11}) (Chan et al., 2014a), with selection possible in as little as a week (Chan et al., 2011). A small number of mAbs currently in the clinic were discovered via phage display, with the most notable being Humira (Adalimumab), currently the world's best selling drug (Urquhart, 2020b), which is used to treat a variety of autoimmune diseases (Broeder et al., 2002).

Phage display was pioneered by George Smith in 1985, where foreign DNA was introduced into phages, such that foreign peptides were fused to the pIII coat protein of bacteriophage M13, resulting in the display of these peptides on the phage surface (Smith, 1985). He demonstrated that it was possible to enrich more than 1000-fold for binders over WT phage, by utilising a specific antibody against the displayed peptide. It was later demonstrated by three independent groups that single-chain variable fragments (scFvs), fusion proteins of the variable regions of the heavy and light chains of antibodies, or Fab fragments, could be presented on phages, permitting the identification of potent, fully human mAb binders (Barbas et al., 1991; Breitling et al., 1991; McCafferty et al., 1990). Since then, alternative antibody formats have been utilised in the construction of antibody phage libraries, including heavy-domain human antibody fragments (V_Hs), heavy-domain camelid and shark antibody fragments (V_HHs) and bivalent scFvs (diabodies) (Abraham et al., 1996; Dooley et al., 2003; Lee et al., 2007).

This approach draws on the possibility to directly link a protein (phenotype) to its associated gene (genotype), to permit multiple rounds of selection and amplification, as illustrated in Fig. 1.11. Typically, antigens are immobilised on ELISA plates or on magnetic beads, although phage display has more recently been used against tumour sections to select for antibodies specific to tumour antigens (Larsen et al., 2015). Phages displaying antibodies are incubated with the targets of interest, before non-binders are removed via washing. Binders are then eluted and amplified via infection of *E. coli*, with the produced phages utilised for the next panning cycle (Fig. 1.11). The repeated rounds of selection enable the enrichment of rare antigen-binding clones, selecting for high affinity binders (Lu

et al., 2020). Concurrently, the short amplification time of phages in *E. coli* enables the identification of mAbs much faster than via traditional hybridoma generation and selection.

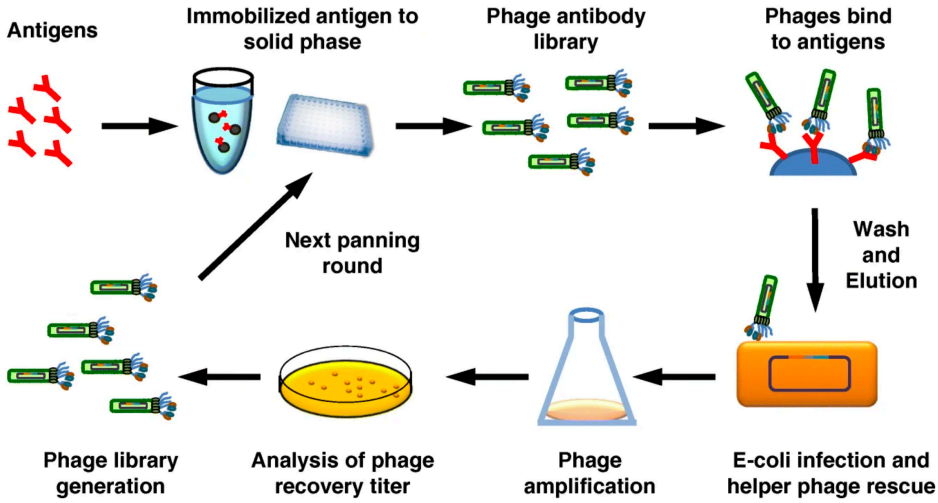


Figure 1.11: Biopanning with an antibody phage-display library. Phages displaying antibodies are incubated with antigens immobilised on a solid surface, such as ELISA plates or magnetic beads. Non-binders are removed via washing, before bound phages are eluted and amplified via infection of *E. coli*. The phages produced can then be utilised for the next panning cycle. After several cycles, the antigen-binding clones are sufficiently enriched and can be further characterised. Figure reprinted from Lu et al., 2020 under the terms of the Creative Commons Attribution License (CC-BY 4.0).

Gene repertoires can be obtained from naïve or immunised animals, or can be synthetically constructed by randomising CDR sequences within fixed frameworks (Lu et al., 2020). Obtaining gene repertoires from naïve or immunised animals involves the reverse transcription and PCR amplification of the relevant segments of the variable heavy (V_H) and light chains (V_L), depending on the library to be constructed. Multiple specific primer pairs are utilised to ensure amplification of different V_H and V_L families. These segments, which represent the Ig gene-encoding repertoire, are ligated into a phage display vector (phagemid), and utilised for phage production (Lu et al., 2020). Care must be taken to maximise diversity of the resulting library. As the natural antibody pairing is not maintained when the phage display libraries are constructed, the resulting antibodies displayed may have suboptimal biophysical characteristics and affinities (Strohl and Strohl, 2012). It must be noted that there are droplet-based approaches to clone paired antibody heavy and light chains into backbones for utilisation in display libraries, to maintain antibody heavy and light chain pairings (Xiao et al., 2017). Similar methods have also previously been mentioned in Sections 1.5 and 1.8.4.

While immunised libraries have antibody genes that have undergone affinity maturation *in vivo*, permitting the development of antibodies with high affinities against the targets, the need for animal models for such libraries, and the need

to generate new libraries for each target limit their widespread use (Lu et al., 2020). In contrast, single naïve and synthetic libraries have been utilised for the identification of high-affinity antibodies against a plethora of targets (Knappik et al., 2000; Vaughan et al., 1996), and most widely accessible commercial libraries are based on non-immunised gene repertoires (Chan et al., 2014a).

Various other types of display technologies have also been employed for the selection of antibodies via directed evolution, including bacterial display (Krüger et al., 2002) and yeast display (Boder and Wittrup, 1997). These technologies have various advantages, such as the fact that they can be used in conjunction with FACS and magnetic assisted cell sorting (MACS). In addition, yeast display utilises eukaryotic post-translational modifications to display scFvs and Fabs, which improves antibody solubility and the processing of complex clones (Sheehan and Marasco, 2015). However, yeast antibody display libraries have a theoretical limit of 10^7 to 10^9 clones in total, which is several orders of magnitude lower than similar phage libraries (Sheehan and Marasco, 2015). In addition, both bacterial and yeast libraries are currently not as well-characterised and optimised as phage libraries.

All the display technologies mentioned merely reveal information on antibody binding and fail to shed any light on antibody functionality, which is a requirement for therapeutic antibody function, where the modulation of target function is key (Seah et al., 2018). In addition, as previously mentioned, antibody display libraries typically do not display complete antibodies, but rather antibody fragments or fusion proteins derived from antibodies. Thus, further genetic engineering is routinely required to convert promising targets into full-length IgG formats. This is occasionally coupled with a loss of binding affinity, and thus, these targets may require further optimisation (Lu et al., 2020).

To overcome these shortcomings, functional antibody screening is frequently carried out via hybridoma technology. The generation of hybridomas via the fusion of B-lymphocytes from mice and immortalised myeloma cells was pioneered by Köhler and Milstein (Köhler and Milstein, 1975), with this technology enabling the isolation and purification of monoclonal antibodies (from a single hybridoma clone) in large quantities.

In order to obtain the necessary B-lymphocytes, mice (or any other mammals) are injected with an antigen that provokes an immune response (Fig. 1.12). Antigens may be accompanied by an adjuvant and multiple injections may be carried out, to enhance the immune response. B-lymphocytes are then extracted and fused with immortalised myeloma cells via electrofusion or the introduction of PEG. Fused cells are selected for by incubation in HAT medium (hypoxanthine-aminopterin-thymidine medium). Aminopterin blocks the pathway that allows for nucleotide synthesis, which results in the death of myeloma cells, which lack the hypoxanthine-guanine phosphoribosyltransferase (HGPRT) gene

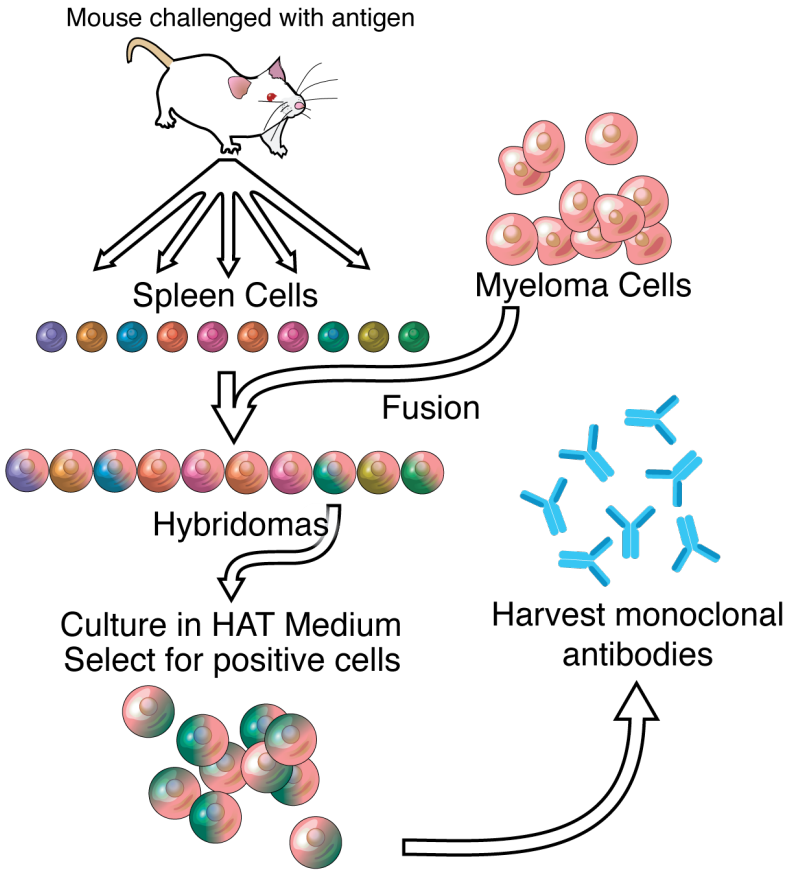


Figure 1.12: Monoclonal antibody generation via hybridoma technology. Mice are immunised with the antigen of interest before extracted splenic B-lymphocytes are fused with myeloma cells to form hybridoma cells. Fused cells are selected for by culture in HAT media. Individual cells of interest can be identified, selected for and clonally amplified, in order to harvest monoclonal antibodies. Figure reprinted from Jones, 2010 under the terms of the Creative Commons Attribution License (CC-BY 3.0).

and thus are unable to produce nucleotides *de novo*. In addition, the unfused B-lymphocytes have a short life span, such that both cell types die in HAT media, with only the B-lymphocyte-myeloma hybrids surviving. These hybridoma cells both produce antibodies and are immortal, with clonal populations of hybridoma cells (those descended from the same parental hybridoma cell) producing monoclonal antibodies targeting the same epitope. The supernatant of each clonal population can then be tested for antibody binding and functionality. Over 90% of antibodies that have been approved by the United States FDA for use in the clinic were generated via hybridoma technology and are used either directly or as chimeric or humanised versions (Parray et al., 2020).

As previously mentioned, fully murine antibodies are limited in their functionality in humans, due to their immunogenicity and rapid clearance in humans (Yamashita et al., 2007). While this has been partially overcome by the

use of chimeric and humanised mAbs, researchers have also established various methods for the discovery and generation of human mAbs. For example, transgenic mouse models have been generated, where the murine Ig-heavy and Ig-light chain loci are disrupted and where transgenes encoding human IgG isotypes are expressed, enabling the application of hybridoma technology to generate human antibodies (Lonberg, 2005).

However, all hybridoma-based screening requires clonal cell expansion to obtain a sufficient antibody concentration for subsequent assays, and thus the requirement for clonal cell expansion, and thus for cell immortalisation makes hybridoma generation time-consuming and expensive, typically limiting the number of clones that can be screened to a few thousand. This constitutes a mere fraction of the antibody repertoire in mammals (approximately $1/10^4$ of the mouse antibody repertoire) (Reddy et al., 2010). Additionally, hybridomas can only be generated from a subset of B-lymphocytes, not including fully matured bone marrow plasma cells, which secrete the vast majority of high affinity IgGs *in vivo* (Reddy et al., 2010). These factors limit the usage of hybridoma technology for the high-throughput screening of therapeutic antibodies.

While alternative immortalisation techniques have been explored, including immortalisation with the Epstein-Barr virus (EBV) (Traggiai et al., 2004) or via genetic reprogramming (Kwakkenbos et al., 2010), these methods remain largely limited to human memory B-lymphocytes and remain time-consuming and limited in throughput.

The above factors propound the benefits of direct screening of non-immortalised B-lymphocytes, which enables the quick isolation of potential mAbs from few cells. As B-lymphocytes express specific cell surface markers depending on their stage of development, primary antibody-secreting cells can be easily isolated via FACS (Kaminski et al., 2012). As certain B-lymphocyte populations, such as memory B-lymphocytes, express surface Ig receptors, they can be incubated with fluorescently-labelled antigens, and screened for binding affinity using FACS (Scheid et al., 2009). Alternatively, B-lymphocytes can be co-cultivated with antigen-coated beads or antigen-expressing cells, and with fluorescence-labelled secondary antibodies. This results in a "halo of fluorescence" being seen in the vicinity of B-lymphocytes expressing antibodies with the binding properties of interest (Tickle et al., 2009). Individual B-lymphocytes can then be picked with a micromanipulator. Once individual B-lymphocytes of interest have been isolated, whether by FACS or direct cell picking, cDNA can be prepared from the cells via reverse transcription and amplification of the heavy and light chains. As previously mentioned, this is not trivial, given the variability in IgH and IgL leader sequences. Thus, a primer-set mixture typically has to be prepared and optimised, to cover different leader sequences (forward primers) and Ig constant regions (reverse primers), to improve recovery (Tiller et al., 2008). These genes can then be cloned

into mammalian expression cell lines, to generate recombinant mAbs for further characterisation (Tiller et al., 2008).

1.9.2 Antibody discovery with microfluidics

Microfluidics provides various advantages over bulk systems in the discovery and study of antibodies. The largest advantage is the fact that microfluidics permits small assay volumes from femtolitre to nanolitre scales, which enable the accumulation of detectable concentrations of antibodies from single antibody-secreting cells in the matter of minutes. For example, antibody secretion rates of single mouse hybridoma cells in 50 pL droplets were measured to be around 4 x 10^4 molecules per cell per minute, resulting in detectable concentration of antibodies (approximately 20 nM) after merely 15 minutes (Mazutis et al., 2013). This facilitates the direct screening of single antibody-secreting cells and eliminates the need for cell immortalisation, which paves the way for faster and cheaper isolation of antibodies of interest.

Various microfluidic and nanowell systems have been utilised to study antibody binding and identify antibodies of interest. Singhal *et al.* have described a bead-based technology to measure antibody-antigen binding kinetics. The technology utilises valves for reagent addition and washing steps, together with beads coated with Protein A that enables antibody capture. Subsequent addition of fluorescently-labelled and unlabelled antigen enabled the measurement of binding kinetics, with a sensitivity comparable to surface plasmon resonance (SPR), but with a sample consumption 4 orders of magnitude lower (Singhal et al., 2010). While valve-based systems are limited in their throughput, a very high level of control is possible, enabling complex reagent addition and washing steps.

In contrast, nanowells enable the high-throughput interrogation of antibody-secreting cells. Jin *et al.* carried out an antibody binding screen by coating nanowells with anti-immunoglobulin antibodies, enabling the capture of antibodies secreted by single B-lymphocytes seeded in the nanowells. Fluorescently-labelled antigen was then added, and antigen binding formed distinctive circular spots, allowing for easy identification of cells of interest, which could then be retrieved with a micromanipulator and subjected to sequencing and cloning (Jin et al., 2009).

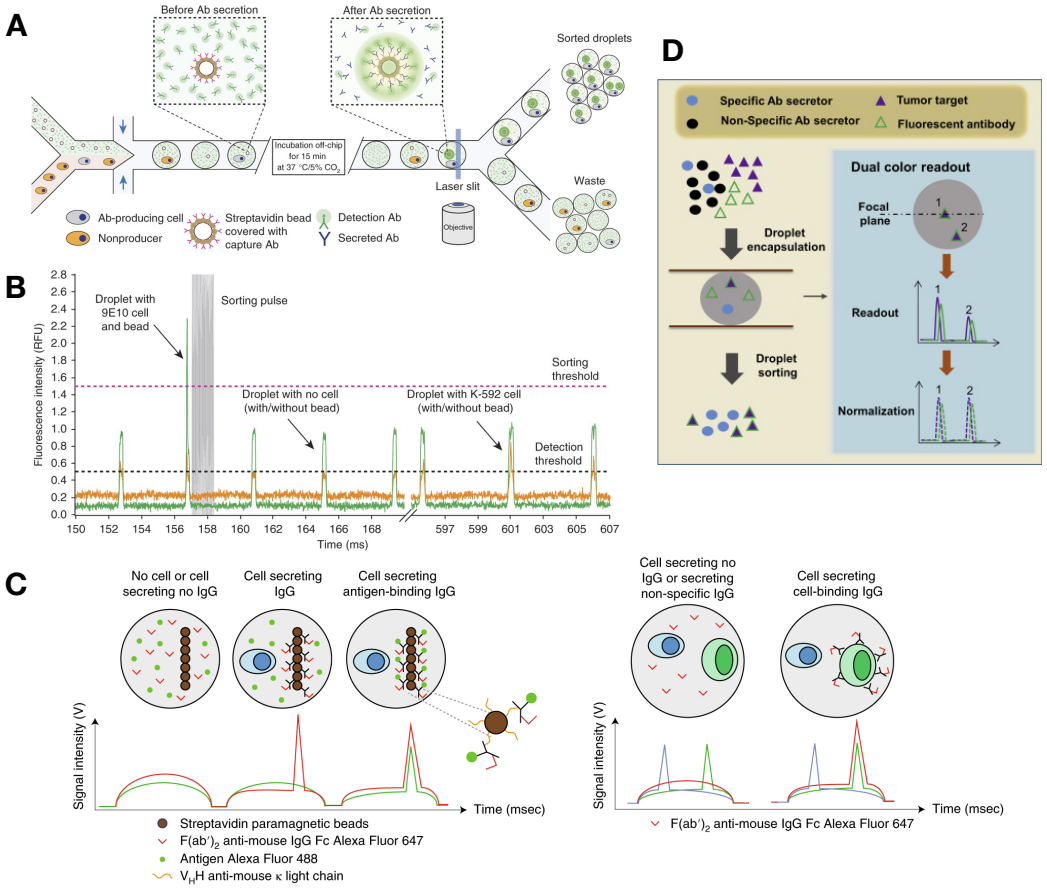
Similarly, the secreted antibodies can be captured on glass slides coated with anti-immunoglobulin antibodies that are placed on top of nanowells. These glass slides can be easily rinsed and stained with fluorescent antigens, while the cells continue to be cultured in the nanowells, or expanded or utilised for further analyses (Love et al., 2006; Ogunniyi et al., 2009). In addition, similar assays have been used to collect multiparametric datasets that study antibody specificity, isotypes and affinity (Story et al., 2008). These nanowell technologies enable fast and high-throughput studies of antibody binding, permitting the study of hundreds of thousands of antibody secreting cells within hours.

While the study of antibody secreting cells has been carried out in single-phase and valve-based microfluidic systems, the advantages of droplet microfluidics make it ideal for the binding and functional screening of antibodies. In particular, the extremely high-throughputs that can be achieved via droplet production theoretically enable screening of a large diversity of cells. In addition, the ease and possibility of droplet manipulation, including sorting with fluorescence-activated droplet sorting (FADS) and adding reagents via droplet fusion and picoinjection, makes droplet microfluidics advantageous for complex multi-step assays, including functional antibody screening assays.

1.9.3 Antibody binding screening with droplet microfluidics

Droplet microfluidics has been utilised for antibody binding screening, in a manner similar to fluorometric microvolume assay technology (FMAT) (Lee et al., 2008). Mazutis *et al.* sorted out antibody-secreting cells from non-secretors in a proof of principle experiment (Fig. 1.13A) (Mazutis et al., 2013). Cells were co-encapsulated with antibody-capturing beads and fluorescently-labelled secondary antibodies. In droplets containing antibody-secreting cells, the secreted antibodies bound to the beads, resulting in the localisation of the secondary antibodies, and thus of their corresponding fluorophores, to the beads. These droplets induced high intensity green fluorescence peaks when they passed the detector (Fig. 1.13B). In contrast, the secondary antibodies remained homogeneously distributed in droplets containing non-secretors, resulting in broad, low intensity peaks (Fig. 1.13B). Cells were stained with a different fluorophore, which enabled the detection of droplets containing cells. Droplets were sorted at a frequency of approximately 200 Hz, enabling the screening of thousands of cells screened with hours of microfluidic operation.

A similar technology with the same principles has also been developed, which combines droplet microfluidics and imaging to permit massively parallel kinetic analyses of antibody-secreting single cells, encompassing simultaneous measurements of antibody secretion rate, specificity and antigen affinity (Eyer et al., 2017). DropMap utilises paramagnetic nanoparticles instead of a single bead and these nanoparticles can be induced to form an elongated aggregate, known as a beadline, via the introduction of a magnetic field. This increases the antibody binding capacity and reduces the number of droplets without beads. The nanoparticles are coated with anti-mouse Igκ nanobodies to capture secreted antibodies, and are encapsulated together with an antibody-secreting cell, Alexa 647-labelled F(ab')$_2$ specific for mouse IgG Fc and Alexa 488-labelled antigen. Secreted IgG and the anti-IgG(Fc) can then be captured onto the beadline, with the ratio of red fluorescence on the beadline relative to that within the droplet used to calculate the IgG concentration (Fig. 1.13C, left). Similarly, if the secreted IgG binds the antigen, the ratio of green fluorescence on the beadline relative to that within the droplet can be used to calculated the strength of the interaction (K_d). DropMap



(see next page)

was used to characterise over 16,000 primary mouse primary B-lymphocytes in tens of thousands of droplets in a single array. Over one million droplets (i.e., 200,000 to 400,000 individual cells) can be analysed in a single 46 x 46 mm chamber, demonstrating the scalability of the technique.

While the detection of antigen binding to beads has been enhanced by the use of the beadline, detecting the binding of antibodies to cells and membrane-bound target is further complicated by the fact that cell-based fluorescence signals are highly variable based on the position of the fluorescent object within the droplet. For example, the closer a cell is to the focal plane, and to the center of the laser spot, the higher the detected fluorescence intensities. However, as these variations in intensities are consistent across different fluorescence channels, one can overcome this confounding effect by the use of a dual-colour normalised readout (Fig. 1.13D) (Shembekar et al., 2018). When a bead or cell is outside the focal plane, both the assay signal and the marker signal show reduced intensity, permitting straightforward normalisation. This has been utilised by Shembekar *et al.* for a high-throughput droplet microfluidic single-cell screening for cells secreting antibodies that bind to target cells, with 220-fold enrichment of cells secreting binding antibodies and 80,000 clones sorted in a single experiment.

Figure 1.13: Screening for antibody binding with droplet microfluidics. **A.** Illustration of the sorting strategy used by Mazutis *et al.*. Beads (hollow circles) and cells (grey or orange ellipses) are co-encapsulated and incubated. Secreted antibodies (dark blue) are captured by capture antibodies (pink) on the streptavidin beads. This results in the localisation of detection antibodies (green) around the bead. However, droplets with no bead, no cell, or a cell that does not secrete antibody, will not present this localised green fluorescence. Droplets can then be sorted by FADS, to selectively enrich for droplets with cells secreting antibodies of interest. Figure adapted from Mazutis et al., 2013 with permission from Springer Nature. **B.** Fluorescence intensities of droplets during analysis and sorting. Cells are stained with orange dye and can be identified by the orange peaks that go over the detection threshold. The green peaks represent the secondary antibodies present. The presence of a high green peak indicates that antibody is secreted and captured by the bead. This triggers sorting when it goes over the sorting threshold (grey area, sorting pulse). Figure reprinted from Mazutis et al., 2013 with permission from Springer Nature. **C.** Illustration of the antibody binding assay used in Eyer et al., 2017 and Gérard et al., 2020. To screen for antibodies against soluble antigens, antibody-secreting cells are encapsulated with streptavidin paramagnetic beads, $F(ab')_2$ anti-mouse IgG Fc - Alexa Fluor 647 and antigen - Alexa Fluor 488 (left). IgGs secreted by antibody-secreting cells are captured on the beads, and will bind anti-mouse IgG Fc Alexa Fluor 647, giving a red fluorescence peak. If the IgG binds the antigen of interest, the antigen - Alexa Fluor 488 will also localise at the beadline, giving a green fluorescence peak. Screening for cell-binding antibodies requires the co-encapsulation of antibody-secreting cells (blue) with target cells (green) and $F(ab')_2$ anti-mouse IgG Fc - Alexa Fluor 647 (right). If the IgG binds the antigen of interest, the antibodies and the anti-mouse IgG Fc Alexa Fluor 647 will localise around the cell, resulting in the presence of a red fluorescence peak. Figure reprinted from Gérard et al., 2020 with permission from Springer Nature. **D.** Distinguishing between antibody-secreting cells that secrete binders and non-binders. Antibody-secreting cells (circles) are co-encapsulated with target cells (purple triangles) and fluorescent secondary antibody (hollow green triangles). If the secreted antibodies bind to the target cells, a sharp fluorescent peak will be observed around the cell. A dual colour fluorescence readout is used to normalise the antibody signal for the confounding factor of cell position within the droplet. Figure modified and reprinted from Shembekar et al., 2018 under the terms of the Creative Commons Attribution License (CC-BY 4.0).

The DropMap technology has been further enhanced to enable the screening of both soluble and membrane-bound antigens, and has been combined with antibody V-gene sequencing via in-droplet single-cell barcoded reverse transcription, in a system termed CelliGO (Gérard et al., 2020). For the detection of antibody-secreting cells secreting antibodies against the soluble targets, droplets similar to those described for DropMap were utilised (Fig. 1.13C, left). In contrast, for the detection of binding to membrane-bound targets, antibody-secreting cells were co-encapsulated with target cells, in the presence of Alexa 647-labelled $F(ab')_2$ specific for mouse IgG Fc (Fig. 1.13C, right). Localisation of the Alexa 647 signal to the cells indicated antibody binding to the target cells. The generated droplets were then sorted via FADS at a frequency of up to 600 Hz, which is equivalent to the sorting of 120 to 240 cells per second, given encapsulation at λ = 0.2 - 0.4, with a total of 1.5 - 11 million cells screened per experiment. The sorted cells were then recovered and re-encapsulated into droplets containing lysis buffer, reverse transcriptase and

hydrogel beads carrying barcoded primers for the capture of antibody heavy and light chain mRNAs. In a manner similar to outlined in Section 1.8.5, the barcodes present on each bead are bead-specific, such that cDNAs generated from each cell are differentially barcoded, enabling cognate V_H and V_L pairs to be identified after NGS. Of the antibodies generated from the hit sequences, 93% and 14% recognised the soluble and membrane-bound antigens of interest, respectively. More non-redundant IgGs were identified as would have been from traditional non-microfluidic antibody screening methods, and the IgGs identified had higher affinities.

The authors also utilised the technology for concurrent targeted single-cell RNA-seq of 32 different transcripts to study cell phenotype, in a manner similar to the targeted single-cell transcriptomic analyses described earlier in Section 1.6.6.

1.9.4 Functional antibody screening in droplet microfluidics

Antibody screening to identify neutralising antibodies against infectious agents is theoretically possible. Wippold *et al.* have carried out a proof of concept experiment to demonstrate that droplet microfluidics can be utilised to identify droplets containing hybridoma cells secreting antibodies neutralising viral infection. They showed this with the in-droplet infection of rat pulmonary epithelial L2 cells by replication-competent GFP-tagged murine hepatitis virus (MHV), and demonstrated that the neutralisation of viral infection by neutralising antibodies could be detected via a fluorescence readout (Wippold et al., 2020). While the imaging readout that was used is not compatible with high-throughput screens, one could imagine that similar assays could be carried out with fluorescence-activated droplet sorting (FADS), as was previously described (Gérard et al., 2020; Shembekar et al., 2018), to enable the high-throughput screening and identification of antibodies that neutralise infectious agents.

High-throughput antibody screening for functional antibodies against a soluble target, the congestive heart failure drug target angiotensin-converting enzyme 1 (ACE-1), was carried out using droplet microfluidics (Fig. 1.14) (El Debs et al., 2012). A heterogeneous hybridoma cell population in which only one in ten thousand cells secreted an antibody inhibiting ACE-1 was encapsulated together with ACE-1, and incubated off-chip to permit antibody accumulation. These droplets were then re-injected into a microfluidic device and fused with a second droplet species containing all reagents required for a fluorescence assay for ACE-1 activity. Droplets without ACE-1 inhibition had high ACE-1 activity and thus high fluorescence intensity, while droplets with ACE-1 inhibition exhibited low fluorescence intensity (Fig. 1.14A). The latter category of droplets were selectively collected by FADS and the encapsulated cells were recovered (Fig. 1.14B). Over 90% of these cells expressed the antibody inhibiting ACE-1, demonstrating that this technology is capable of a more than 9000-fold enrichment of cells secreting a functional antibody. Over

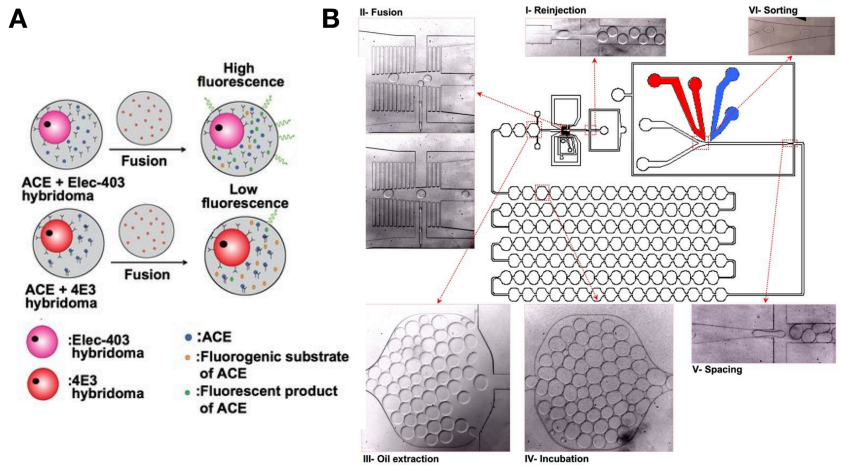


Figure 1.14: Experimental setup for functional antibody screening. **A.** Two different hybridoma cell lines, which expressed either the ACE-1 inhibitory antibody 4E3 or the non-inhibitory antibody Elec-403, were encapsulated into droplets together with recombinant ACE-1. The droplets were incubated to permit antibody accumulation, and then fused with reagents for a fluorogenic assay for ACE-1 activity. Droplets that contained the Elec-403 hybridoma, which secreted a non-inhibitory antibody, had uninhibited ACE-1, leading to high fluorescence readouts in the ACE-1 enzymatic assay. In contrast, droplets that contained the 4E3 hybridoma demonstrated reduced ACE-1 activity, and thus had a low fluorescence readout. **B.** Microfluidic chip utilised for the reinjection of droplets after a 6 h incubation off-chip (I). Droplets were fused with droplets containing the fluorogenic ACE-1 substrate (II) and incubated for 30 min in a delay line (IV). The sorting module (VI) permitted the selective collection of droplets with low fluorescence intensities, which presumably contained inhibited ACE-1. Figures adapted from El Debs et al., 2012. Copyright 2012 National Academy of Sciences.

300,000 cells were processed in a single experiment, which took less than a day, demonstrating the speed and throughput possible with this technology. While hybridoma cells were utilised here, such technology should also be applicable to non-immortalised B-lymphocytes, as was demonstrated in Eyer et al., 2017 and Gérard et al., 2020.

Droplet microfluidic methods for antibody screening for binding or function mostly require prior knowledge or predetermined targets, making them less useful for target-agnostic screening. In addition, functional antibody screening in droplets generally requires assay systems utilising fluorescence, which necessitates the availability of compatible fluorescence assays, which may not exist or may require prior development.

Current bulk antibody screening technologies are either limited to screening for antibody binders (phage display) or are able to assess functionality, but are time-consuming, expensive and limited in the diversity of clones that can be obtained and screened (hybridoma technology) (as outlined in Section 1.9.1). Droplet microfluidics has been used to overcome these limitations, with functional single-cell hybridoma screening having been demonstrated for soluble targets (as exemplified in Section 1.9.4) (El Debs et al., 2012), but this has not yet been demonstrated

for membrane-based targets or whole cells. In addition, most antibody discovery platforms currently require prior knowledge about the target, such that screening cannot take place in a target-agnostic manner. Combining the utilisation of droplet microfluidics in functional antibody screening, with the utilisation of droplet microfluidics for single-cell transcriptomic analyses (as outlined in Section 1.6.4) will enable the development of a method that permits high-throughput target-agnostic functional antibody screening. Such a technology would permit the study of the effect of antibodies on transcriptomes of single target cells, and demonstrating the feasibility of such a technology is one of the main goals of this thesis.

2 Aims and objectives

The main aims of this thesis were to assess the feasibility of a technology combining the uses of droplet microfluidics for single-cell transcriptomic analyses (as outlined in Section 1.6.4) and in functional antibody screening (as explained in Section 1.9.4). This method would enable high-throughput target-agnostic functional antibody screening by using changes in target cell transcriptomes as a readout.

In order to achieve this, we have made preparations for actual screens, by arranging for the production of a diverse hybridoma library secreting antibodies against SK-BR-3 cells, as outlined in **Chapter 3**.

The first important aim was to assess the feasibility of using single-cell transcriptomic analyses to study cell-cell interactions, which is discussed in **Part I**. This encompasses the first objective, which was to establish a Drop-seq workflow to effectively generate and process droplets containing two cell types, which is outlined in **Chapter 4**. In addition, the second objective was to identify suitable model systems for the establishment of microfluidic and sequencing workflows, which is discussed in **Chapter 5**. The third objective was to assess how many single cells would be required to detect a given perturbation, and if a perturbation can be identified from its effect on the transcriptome of a single cell, which is examined in **Chapter 6**. A discussion pertaining to the objectives discussed in **Part I** can be found in **Chapter 7**.

The second important aim was to establish an antibody sequencing technology that would be compatible with Drop-seq. Such a technology would ideally be capable of sequencing and pairing the heavy and light chain variable regions of antibody genes, while ensuring that these sequences can be paired to relevant target cell transcriptomes. The method developed and the associated results are discussed in **Part II**, with the microfluidic pipeline outlined in **Chapter 8**. We analysed both predefined hybridoma mixtures with known antibody heavy and light chain sequences and a complex hybridoma mixture with unknown antibody heavy and light chain sequences, which are presented in **Chapters 9 and 10** respectively. A discussion relevant to **Part II** can be found in **Chapter 11**.

The overall feasibility of a high-throughput target-agnostic functional antibody screening technology and the associated implications are discussed in **Chapter 12**.

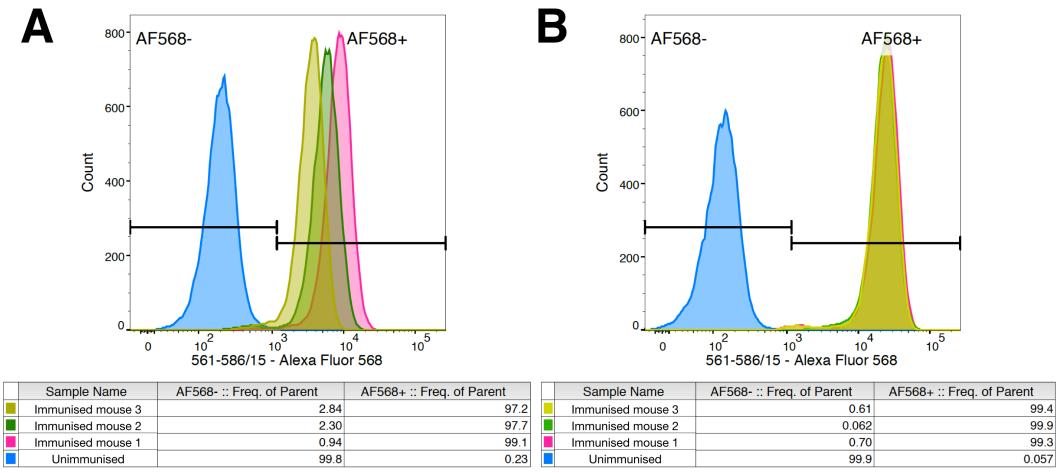
3 Preparation of a complex hybridoma library

For any eventual screens, the structure of which is outlined in Section 12.1, we would require a complex population of antibody-secreting cells, where the different antibody-secreting cells would secrete different antibodies that bind a specific cell line. Therefore, we employed Precision Antibody, a company involved in antibody generation, to generate a diverse hybridoma library expressing different antibodies against SK-BR-3 cells.

The SK-BR-3 cell line is a human immortalised breast cancer cell line that overexpresses the human epidermal growth factor receptor 2 (HER2/ERBB2). This cell line is medically relevant as approximately 20% of breast cancers demonstrate HER2 overexpression, and treatments that specifically target Her2 have been shown to be promising. One such drug, trastuzumab, is a humanised monoclonal antibody against Her2 (Henjes et al., 2012).

Mice were first immunised with whole SK-BR-3 cells, and sera were extracted from the immunised mice. Control immunisation experiments were run in parallel at EMBL and an immune response against SK-BR-3 was detected for both the in-house sera (Fig. 3.1A) and those obtained from Precision Antibody (Fig. 3.1B). Primary B-cells were then extracted from the immunised mice and fused with myeloma cells to generate a diverse population of hybridoma cells. We expect a subset of hybridoma cells to secrete antibodies that bind to surface proteins on SK-BR-3 cells, and that a subset of these would induce transcriptomic changes, which could then be detected via single-cell transcriptomic sequencing, as is outlined in Chapter 5.

We have utilised these hybridoma cells for a number of experiments, including single-cell RNA-seq of SK-BR-3 cells treated with Precision Antibody hybridoma supernatant, as outlined in Section 5.9. In addition, as a proof of concept, this complex hybridoma mixture has been sequenced with our antibody sequencing technology, as explained in Chapter 10.



Sample Name	AF568- :: Freq. of Parent	AF568+ :: Freq. of Parent
Immunised mouse 3	2.84	97.2
Immunised mouse 2	2.30	97.7
Immunised mouse 1	0.94	99.1
Unimmunised	99.8	0.23

Sample Name	AF568- :: Freq. of Parent	AF568+ :: Freq. of Parent
Immunised mouse 3	0.61	99.4
Immunised mouse 2	0.062	99.9
Immunised mouse 1	0.70	99.3
Unimmunised	99.9	0.057

Figure 3.1: Sera binding to whole SK-BR-3 cells. SK-BR-3 cells were incubated with 100x diluted mouse sera, and subsequently with 1000x diluted Alexa Fluor 568 goat anti-mouse IgG antibody before flow cytometry analysis. Alexa Fluor 568 was detected at 561-586/15, x-axes. **A.** Sera from an unimmunised mouse in EMBL (in-house) (blue) and from three mice immunised in EMBL (in-house) with whole SK-BR-3 cells (pink, green, mustard) were utilised for staining. **B.** Sera from an unimmunised mouse from Precision Antibody (blue) and from three mice immunised by Precision Antibody with whole SK-BR-3 cells (pink, green, mustard) were utilised for staining.

Part I

A single-cell transcriptomic method to map cell-cell interactions

4 A workflow for the transcriptomic analysis of cell-cell interactions

This part describes a novel microfluidic workflow for the analysis of interactions between two cell types via single-cell transcriptomics. Cell-cell interactions play an important role in the development and function of multicellular organisms, but current technologies for the study of cell-cell interactions between single cells, in particular, those that are utilised for the screening of cell-cell interaction factors, are frequently only capable of assessing one or a few different variables simultaneously, and largely fail to acquire complex transcriptomic data. Various emulsion-based single-cell sequencing technologies, such as Drop-seq (Macosko et al., 2015), InDrop (Klein et al., 2015) and the 10X Genomics 3' Gene Expression platform (Zheng et al., 2017), are capable of detecting complex transcriptomic data from single cells, but these fail to consider the interactions that these cells may partake in.

The microfluidic pipeline for the transcriptomic analysis of cell-cell interactions is presented in this chapter. To overcome the problems of Poisson distribution that accompany the co-encapsulation of two cell types, picoinjection was utilised to selectively inject lysis buffer into droplets with the desired cell occupancies. We have tested various stimulant and target cell pairs, in a bid to identify a suitable model system, which is outlined in Chapter 5. In addition, we have also carried out analyses to determine how many single cells would be required to detect a given perturbation, which is described in Chapter 6.

4.1 Microfluidic pipeline for the transcriptomic analysis of cell-cell interactions

Our two cell types of interest – stimulant cells and target cells – would be co-encapsulated together with Drop-seq beads. The stimulant cells present or secrete factors that bind to cell surface proteins on the target cell, triggering a transcriptomic change in the target cell. The droplets would then be incubated at

37 °C for 4 - 8 hours, to enable antibody binding and the induction of transcriptomic changes. Subsequently, cell lysis reagents would be selectively picoinjected into the droplets of interest, before the emulsion is broken and the library is prepared. From the sequencing data, one expects to obtain two sets of data, namely sequences relating to the stimulant, together with gene expression data of the target cells, which can be correlated using cell barcodes (Fig. 4.1). Based on the target cell transcriptomes, we expect to detect either single specific stimulant cells of interest, or small groups of identical stimulant cells. Please refer to Chapters 6 and 7 for more information regarding the minimum number of cells required to detect a perturbation.

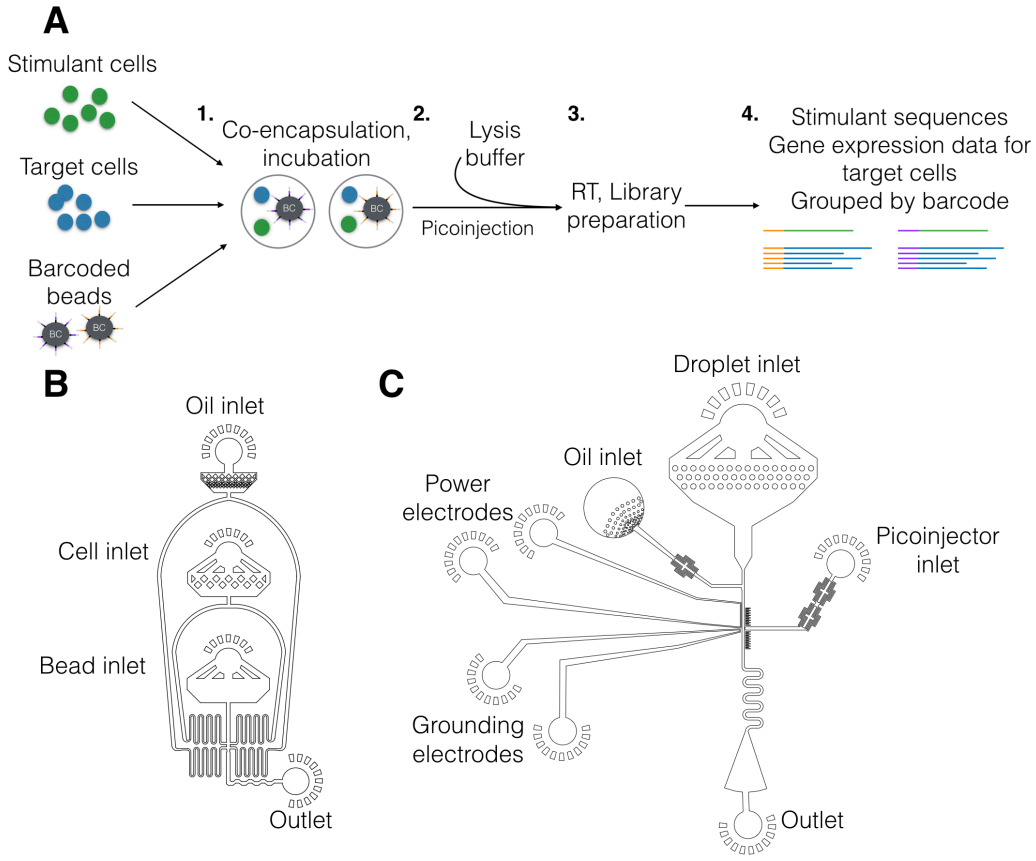


Figure 4.1: Microfluidic pipeline and chip designs for the transcriptomic analysis of cell-cell interactions. A. Microfluidic pipeline for the transcriptomic analysis of cell-cell interactions. **(1)** Stimulant cells are co-encapsulated with target cells and barcoded beads. Droplets are incubated to enable the induction of transcriptomic changes on the target cells. **(2)** Lysis buffer is then added via selective picoinjection, before **(3)** the emulsion is broken and the library is further processed. **(4)** The sequencing data would contain both information on the stimulant (green) and gene expression of the target cells (blue), grouped by barcode (orange, purple). **B.** Chip design used for bead and cell encapsulation (Step 1). Chip design reprinted from Macosko et al., 2015 with permission from Cell. **C.** Chip design used for picoinjection (Step 2). Chip design reproduced with permission from Hongxing Hu, EMBL Heidelberg.

Theoretically, any emulsion-based single-cell transcriptomic technology could be used. However, as we require an incubation step in between cell encapsulation and cell lysis, the cell lysis reagents cannot be added at cell encapsulation, which is the case for typical emulsion-based single-cell transcriptomic workflows. Therefore, we cannot use the 10X Genomics Single Cell 3' Gene Expression technology as the commercial nature of the technology precludes a thorough understanding of the composition of the lysis buffer, preventing us from modifying the workflow. Between Drop-seq and InDrop, Drop-seq has been shown to perform more reliably (Zhang et al., 2019), and the smaller droplets and smaller beads are easier to manipulate and design chips for.

4.2 Co-encapsulation of two cell types

Passive cell encapsulation in droplet microfluidics is non-deterministic and can be best modelled by the Poisson distribution (Clausell-Tormos et al., 2008). In traditional single-cell transcriptomic assays, cells are flowed into the chip at a density of less than one cell per droplet on average, such that a majority of droplets contain no cells. For example, for $\lambda = 0.1$ (where λ is the cell density divided by droplet volume), about 90.5% of droplets would contain no cells, about 9% would contain one cell as desired, while less than 0.5% of droplets would contain two or more cells (Shembekar et al., 2016). Coupled with the high frequency of droplet production, this ensures that a sufficient number of cell-containing droplets is generated for analyses, while minimising the number of droplets containing multiple cells, ensuring true single-cell analyses.

However, Poisson statistics makes the co-encapsulation of two cell types difficult. For the encapsulation of one cell each of two different cell types, as is desired in this experiment, the maximum achievable probability is only 13.5% (Fig. 4.2, dark green) (Hu et al., 2015). Even when considering stimulant cells encapsulated with two or more target cells (Fig. 4.2, light green), which make up 9.71% of droplets and could still yield useful sequencing data, over 60% of droplets containing barcoded beads would contain an undesirable combination of cells (Fig. 4.2, light red, dark red). These droplets would contribute to the sequencing data generated, but as they contain either only one cell type, or more than one stimulant cell, the transcriptomic data generated from these droplets would not be useful for our analysis, and would instead be a waste of sequencing resources.

4.3 Picoinjection

To overcome this problem, we have optimised a technology that could selectively process droplets with the desired cell occupancy, by the selective picoinjection of lysis buffer into droplets with the correct cell occupancy. Our previously encapsulated droplets would be re-injected into the picoinjection chip while the cell lysis reagents

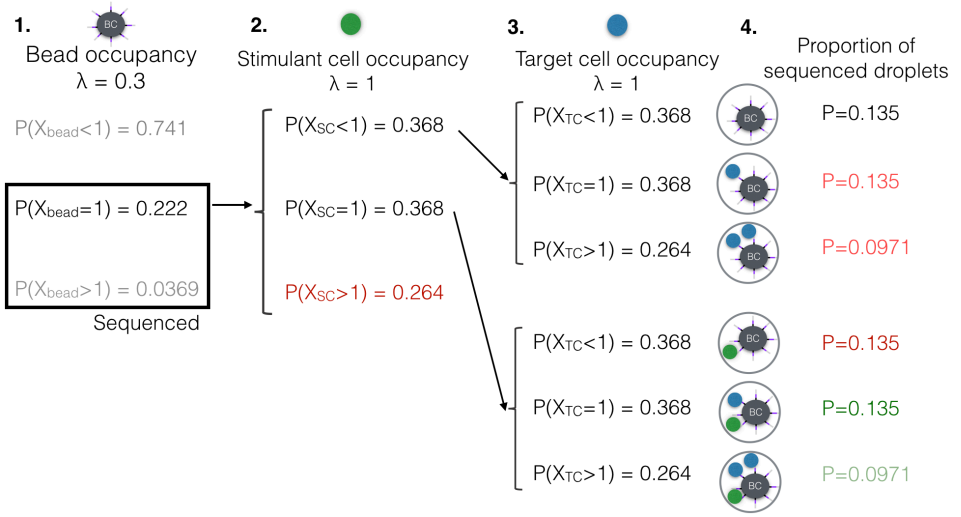


Figure 4.2: Droplet occupancies for encapsulation with two cells and a bead. The proportion of each droplet type are shown for an experiment in which barcoded beads (grey circles, BC) are encapsulated at $\lambda = 0.3$ and the stimulant (green) and target cells (blue) are encapsulated at $\lambda = 1$. **(1)** Approximately 22% of droplets would contain a single bead and 25% of droplets will contain at least one barcoded bead and thus generate sequencing data. **(2)** Of the sequenced droplets, 36.8% contain no stimulant cells, 36.8% contain exactly one stimulant cell, while 26.4% contain two or more stimulant cells. **(3)** Independently, 36.8% of sequenced droplets contain no target cells, 36.8% contain exactly one target cell, while 26.4% contain two or more target cells. **(4)** We can calculate the expected percentages of droplets with different occupancies. We expect that only 13.5% of sequenced droplets will have our desired occupancy of one stimulant cell and one target cell (dark green). Droplets with one stimulant and two or more target cells may still be useful (light green). Droplets containing only target cells (light red) or only stimulant cells or multiple stimulant cells (dark red) will not be useful.

would be loaded into the picoinjector nozzle (Fig. 4.3A). The stimulant and target cells would be stained with different cell dyes (Cell Trace CFSE and Cell Trace Violet) prior to encapsulation. Picoinjection is triggered by the activation of an electric field, which destabilises the surfactant film surrounding droplets and causes injection of reagents into droplets (Abate et al., 2010). We would utilise the lasers of wavelengths 375 nm and 488 nm and their associated photo-multiplier tubes (PMTs) that detect emitted light at 450 nm and 525 nm, and trigger picoinjection only when fluorescence signals from both cell types are detected. This means that lysis buffer will only be added to droplets with the desired occupancy, thus minimising the waste of resources spent on sequencing droplets with undesirable droplet occupancies.

This has been demonstrated for single-colour picoinjection, where reagents were injected into droplets containing green cells (Fig. 4.3A) but not into empty droplets (Fig. 4.3B). Given that dual-colour sorting has been previously established in the lab (Hu et al., 2015), the utilisation of dual-colour picoinjection can be easily envisioned.

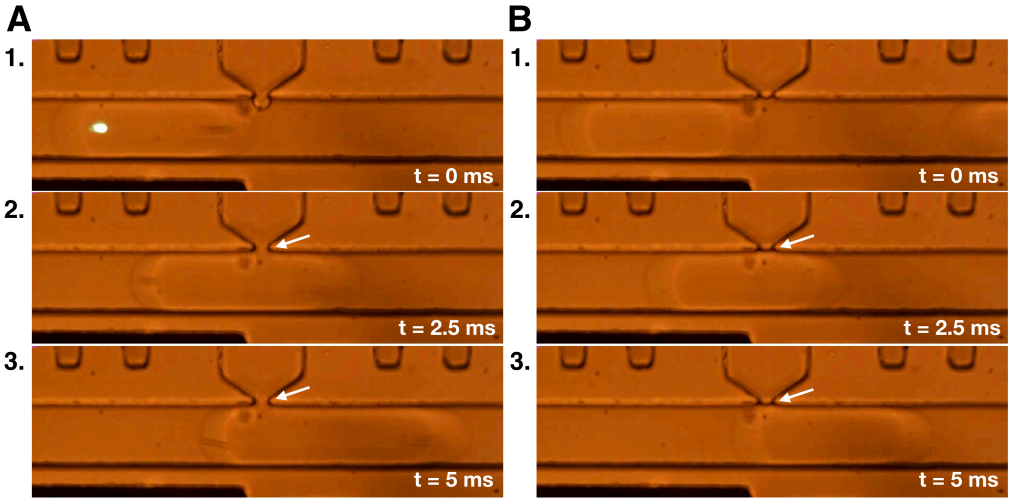


Figure 4.3: Selective processing of droplets with desired cell occupancies via single-colour picoinjection. A. High-speed video frames taken at **(1)** t = 0 ms, **(2)** t = 2.5 ms and **(3)** t = 5 ms, show picoinjection in the presence of one or more green cells in a passing droplet (green spot in (1)). Note the loss of interface between the liquid in the picoinjector and in the droplet (white arrows in (2) and (3)). **B.** High-speed video frames taken at **(1)** t = 0 ms, **(2)** t = 2.5 ms and **(3)** t = 5 ms, show that picoinjection does not take place in the absence of green cells in a passing droplet. Note the presence of the interface between the liquid in the picoinjector and in the droplet (white arrows in (2) and (3)). Frames from a video reproduced with permission from Hongxing Hu, EMBL Heidelberg

5 | Identification of a suitable model system

To establish a technology that can study cell-cell interactions at a single-cell level, we require a suitable model system. This system should be comprised of two cell lines - a stimulant cell line and a target cell line. The stimulant cell line should present or secrete a genetically-encoded factor that would bind to cell surface proteins on the target cell line and trigger a transcriptomic change in the target cell line that would be detectable by single-cell transcriptomics (Fig. 5.1, (1)). Ideally, we would also have a negative cell line that would be identical to the stimulant cell line (Fig. 5.1, (2)), except that it would not present or secrete the stimulant, and thus we expect any target cell co-encapsulated with this cell line to have a transcriptome similar to a cell that was present alone in a droplet (Fig. 5.1, (3)). The stimulant should be genetically encoded, such that it can be detected via emulsion-based single-cell sequencing (Fig. 5.1, (4)). In addition, as both cells would be co-encapsulated in the same droplet and would thus be associated with the same droplet barcode, it would be ideal for the stimulant and target cell lines to be derived from different species, such that the two transcriptomes can be distinguished after sequencing. In addition, the time permitted for the stimulant and cell line of interest should ideally be no more than 8 - 12 hours, due to limitations imposed by the droplet microfluidic format. Whilst it has been shown that cell viability in droplets generally remains above 90% over two days, it has also been demonstrated that increased cell density in droplets results in increased cell death, likely due to the lack of nutrition or the accumulation of toxic metabolites (Clausell-Tormos et al., 2008). Co-encapsulation doubles the effective cell concentration in droplets, and thus we prefer to limit the time the cells spend in droplets to 8 - 12 hours, to minimise cellular stress and cell death.

The stimulant can be either presented on the cell surface or secreted by the stimulant cell line. As the density of the encapsulation medium is still lower than that of cells, the encapsulated cells tend to sink and interact, ensuring transcriptomic changes in the target cells co-encapsulated with a stimulant cell that presents the stimulant on the cell surface (Hu et al., 2015; Segaliny et al., 2018). In addition, the small droplet volumes ensure that any secreted factors, such as antibodies, will reach a functional concentration in the matter of hours (El Debs et al., 2012;

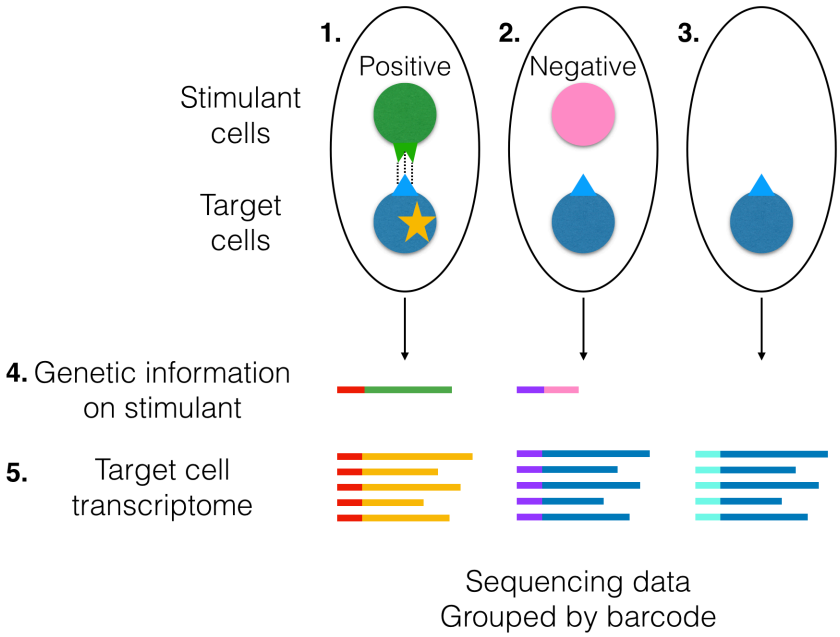


Figure 5.1: Ideal model system for transcriptomic analysis of cell-cell interactions. The ideal model system would comprise of a stimulant cell line (green) and a target cell line (blue). **(1)** Droplet containing a stimulant cell (green) together with a target cell (blue). The stimulant cell line should produce or secrete a factor that binds to and induces transcriptomic changes in the target cells (yellow star). **(2)** Droplet containing a cell that does not express the stimulant, but is otherwise genetically identical to the stimulant cell (pink), together with a target cell (blue). No transcriptomic changes would induced in the target cells (blue). **(3)** Droplet containing only a single target cell (blue). Target cells present alone in a droplet (blue) should have transcriptomes similar to the target cell in (2). **(4)** The stimulant should be genetically encoded such that cells that present and do not present the stimulant can be differentiated via sequencing (green, pink). **(5)** Sequences from the same droplets, namely stimulant-associated sequences and target cell transcriptome sequences, can be linked via cell barcode sequences (red, purple and cyan for (1), (2) and (3) respectively). The target cell transcriptomes should be distinguishable from the stimulant cell transcriptomes, and those from target cells co-encapsulated with positive stimulant cells (1, yellow) and negative stimulant cells (2, dark blue) should be distinguishable via single-cell sequencing.

Shembekar et al., 2018). This has been exemplified by Mazutis *et al.*, where detectable concentrations of antibodies (approximately 20 nM) were detected after merely 15 minutes (Mazutis et al., 2013).

These three identified cell lines would then be used for our subsequent screens, where stimulant-presenting cells would be spiked into a population of negative cells at different ratios, prior to co-encapsulation with the target cell line, to elucidate the sensitivity and specificity of our technology.

We have identified and tested various model systems. Many of the initial systems tested involved antibody-secreting hybridoma cells together with their associated target cells, as we hoped to couple the study of cell-cell interactions with the antibody sequencing technology developed in Part II, by establishing and

optimising a system to study the effect of antibodies on target cell transcriptomes. Subsequently, other model systems involving interactions via membrane-bound proteins were also considered and tested.

As antibody binding can indeed induce effects on target cell transcriptomes (Abidi et al., 2008; Sousa et al., 2019), we have tested various promising systems, first with bulk sequencing, to assess if there are indeed transcriptomic differences between treated and untreated populations. Systems have then been subjected to emulsion-based single-cell sequencing, to assess if the transcriptomic differences can be identified at a single-cell level. The 10X Genomics Chromium platform for 3' Gene Expression is currently the gold standard in the field and was utilised, in conjunction with cell hashing (explained in Section 5.7) to minimise batch effects.

For all experiments involving hybridoma cells, the H25B10 mouse hybridoma cell line was used as a negative control. H25B10 hybridoma cells secrete antibodies against the Hepatitis B virus Surface Antigen, which is not present on the target cell lines utilised.

5.1 SK-BR-3 cells with 4D5 hybridoma cells

The SK-BR-3 cell line is a human immortalised breast cancer cell line that overexpresses the human epidermal growth factor receptor 2 (HER2/ERBB2) and is thus susceptible to trastuzumab, a humanised monoclonal antibody against Her2 (Henjes et al., 2012). The murine version of this antibody, 4D5, is secreted by the 4D5 hybridoma cell line. As the effect of the anti-Her2 antibodies on the SK-BR-3 transcriptome is well defined (Le et al., 2005), we expected that SK-BR-3, together with trastuzumab and the associated 4D5 hybridoma cell line, would be a good model system for us to optimise and study the sensitivity of our assay.

5.1.1 4D5 hybridoma cells fail to produce functional antibody

In order to accurately estimate antibody concentrations within hybridoma supernatants, we carried out an ELISA against mouse IgG (Fig. 5.2A). However, we estimated the soluble IgG levels in 4D5 hybridoma supernatants to be between 74.0 ng/ml and 87.2 ng/ml (Fig. 5.2A, undiluted (green crosses) and 4x diluted (mustard crosses)), which is significantly lower than the antibody concentrations of other hybridoma supernatants. For example, the mouse IgG concentrations of the 8G5F11, IE9F9 and H25B10 hybridoma supernatants were found to be 7.74 µg/ml, 12.09 µg/ml and 24.72 µg/ml respectively (Fig. 5.2B). This makes the antibody concentration of the 4D5 hybridoma supernatant over 85x lower than expected.

It is also possible to detect antibodies on the surface of hybridoma cells by staining hybridoma cells with antibodies against the IgGs of interest. While the IgG concentrations present on the cell surface may not necessarily correlate well with

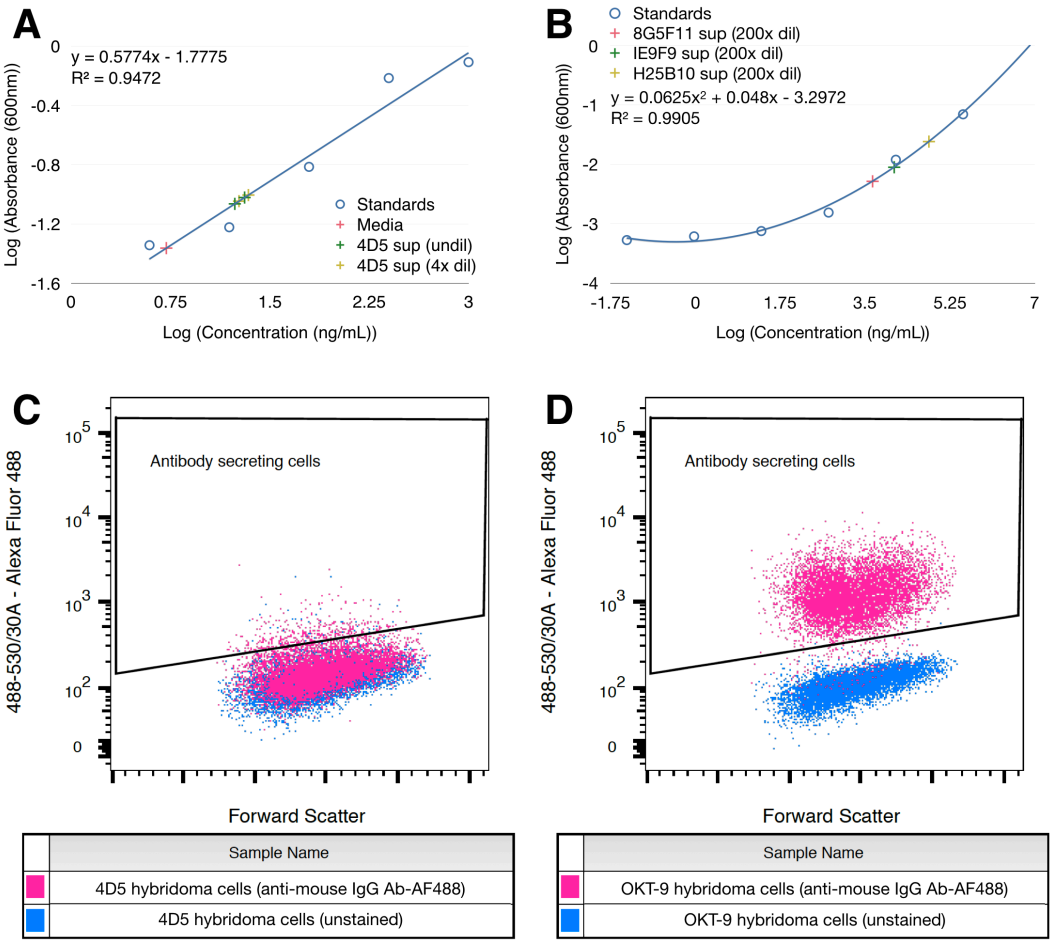


Figure 5.2: 4D5 hybridoma cells express low antibody amounts. A. The mouse IgG concentrations of hybridoma supernatants were assessed by ELISA. A standard curve was prepared (blue line) based on the absorbance values at 600 nm obtained for mouse IgG standards (hollow blue circles). The common logarithms of the absorbance values at 600 nm are shown on the y-axis while the common logarithms of mouse IgG concentrations in ng/ml are shown on the x-axis. Mouse IgG concentrations of undiluted 4D5 (green crosses) and 4x diluted 4D5 supernatants (mustard crosses) were assessed. Media was also assessed as a negative control (pink cross). Data was generated with the IgG mouse ELISA Kit from abcam. **B.** The mouse IgG concentrations of hybridoma supernatants were assessed by ELISA. A standard curve was prepared (blue line) based on the absorbance values at 600 nm obtained for mouse IgG standards (hollow blue circles). The natural logarithms of the absorbance values at 600 nm are shown on the y-axis while the natural logarithms of mouse IgG concentrations in ng/ml are shown on the x-axis. Mouse IgG concentrations of 200x diluted 8G5F11 (pink cross), 200x diluted IE9F9 (green cross) and 200x diluted H25B10 supernatants (mustard cross) were assessed. Data was generated with the IgG mouse ELISA Kit from abcam. **C.** 4D5 hybridoma cells were incubated with no antibody (blue) or with 2000x diluted Alexa Fluor 488 goat anti-mouse IgG antibody (pink) before flow cytometry analysis. **D.** OKT-9 hybridoma cells were incubated with no antibody (blue) or with 2000x diluted Alexa Fluor 488 goat anti-mouse IgG antibody (pink) before flow cytometry analysis. Alexa Fluor 488 was detected at 488-530/30, y-axes and forward scatter is reflected on the x-axes.

secreted IgG antibodies, it has been shown that hybridoma cells that secrete IgG generally also had surface-bound IgG (Liu et al., 2015a).

We found that there was a negligible amount of surface-bound IgG present on 4D5 hybridoma cells (Fig. 5.2C), as indicated by the low Alexa Fluor 488 signal on stained 4D5 hybridoma cells. In contrast, the high Alexa Fluor 488 signals present on OKT-9 hybridoma cells, indicate high expression of surface-bound IgG (Fig. 5.2D).

These data suggest that 4D5 hybridoma cells secrete a limited amount of antibody and thus, may be restricted in their utility as a model system.

In addition, ten 4D5 heavy chain sequences were analysed with IgBLAST (Ye et al., 2013), revealing that while a majority of sequences (9/10) align well to each other and have identifiable CDR3 regions, the sequenced heavy chains are non-productive (Fig. 5.3).

Query= 4D5 Heavy Chain

Length=428

Sequences producing significant alignments:	Score (Bits)	E Value
IGHV9-1*02germline gene	195	1e-51
IGHV9-3-1*01germline gene	191	1e-50
IGHV9-2*02germline gene	188	9e-50
IGHD2-10*01germline gene	21.8	0.033
IGHD2-11*02germline gene	21.8	0.033
IGHD2-4*01germline gene	21.8	0.033
IGHJ4*01germline gene	69.9	3e-16
IGHJ2*02germline gene	27.6	0.002
IGHJ3*01germline gene	25.7	0.006

Domain classification requested: imgt

V-(D)-J rearrangement summary for query sequence (multiple equivalent top matches, if present, are separated by a comma):

Top V gene match	Top D gene match	Top J gene match	Chain type	stop codon	V-J frame	Productive	Strand
IGHV9-1*02	IGHD2-10*01,IGHD2-11*02,IGHD2-4*01	IGHJ4*01	VH	No	Out-of-frame	No	+

V-(D)-J junction details based on top germline gene matches:

V region end	V-D junction*	D region	D-J junction*	J region start
CTGTG	ATAGG	CCTACTATGATAAC	GGAGGG	ATGCT

*: Overlapping nucleotides may exist at V-D-J junction (i.e, nucleotides that could be assigned to either rearranging gene). Such nucleotides are indicated inside a parenthesis (i.e., (TACAT)) but are not included under the V, D or J gene itself.

Sub-region sequence details:

	Nucleotide sequence	Translation	Start	End
CDR3	GATAGGCCTACTATGATAACGGAGGGATGCTATGGCCTAC	DRPTMITEGCYGL	318	357

Alignment summary between query and top germline V gene hit:

	from	to	length	matches	mismatches	gaps	identity(%)
FR2-IMGT	177	179	3	3	0	0	100
CDR2-IMGT	180	203	24	22	2	0	91.7
FR3-IMGT	204	317	114	107	7	0	93.9
CDR3-IMGT (germline)	318	318	1	1	0	0	100
Total			142	133	9	0	93.7

Figure 5.3: 4D5 heavy chain is non-productive. Ten PCR fragments were amplified from 4D5 cDNA and sent for sequencing. 9/10 sequences were homologous and this sequence was subjected to IgBLAST (Ye et al., 2013), which indicated that this sequence was non-productive.

The lack of a functional heavy chain detected from 4D5 hybridomas, accompanied by the low antibody concentrations in 4D5 hybridoma supernatants suggest that limited, if any, functional antibody is produced by the 4D5 hybridoma cells. The cell line was also discontinued in ATCC in 2018, with no indication of if, or when, it would be reinstated. Therefore, it is not feasible to use 4D5 as part of a model system for the establishment of our technology.

5.1.2 Single-cell RNA-seq via Drop-seq

SK-BR-3 cells with anti-Her-2 antibody

We also assessed the transcriptomic responses of SK-BR-3 cells encapsulated in the absence or presence of 1 µg/ml recombinant anti-Her2 antibody. Clustering by sample is observed for the differentially treated SK-BR-3 cells along the second component, which is responsible for 6% of the variance (Fig. 5.4).

This suggests that the transcriptomic changes induced by this stimulus are detectable by single-cell transcriptomics. However, the lack of 4D5 cells expressing functional anti-Her-2 antibodies currently prevents us from developing this model system further.

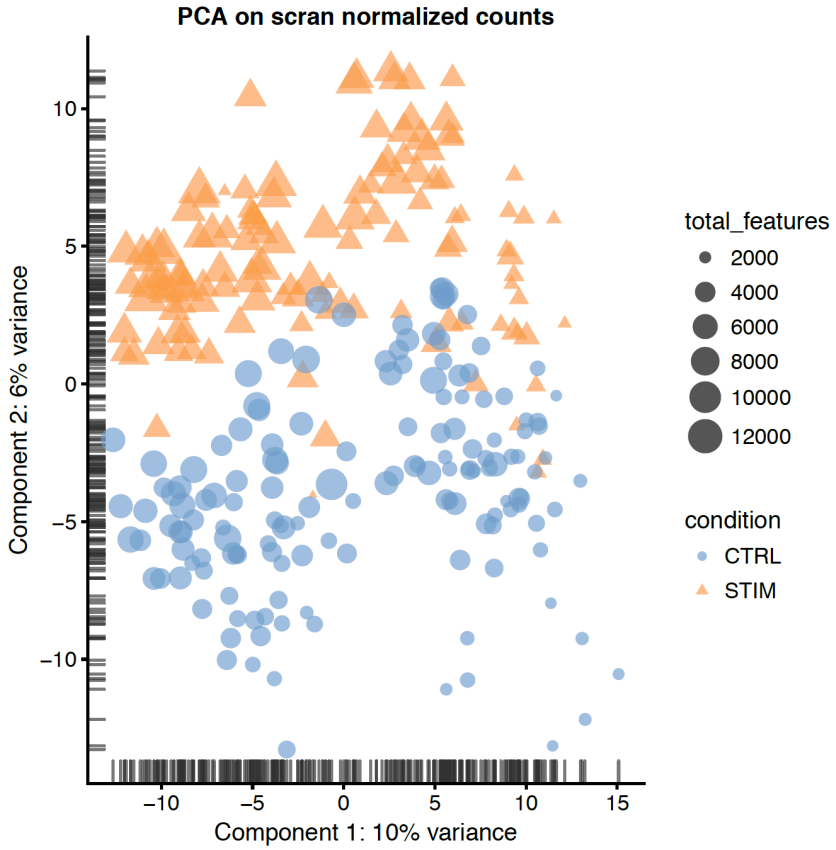


Figure 5.4: Dimensional reduction of gene expression data of SK-BR-3 cells. Gene expression data of untreated SK-BR-3 cells (blue circles) and SK-BR-3 cells treated with 1 µg/ml recombinant anti-Her2 antibody (orange triangles), obtained by Drop-seq. The Principal Component Analysis was carried out on the scran normalised counts, with the first two principal components plotted. The size of the shapes indicate the number of total features, namely the number of genes that are detected per droplet. Data reproduced with the permission of Charles Girardot, EMBL Heidelberg.

5.2 Jurkat cells with hybridoma cells secreting anti-CD4 antibodies

The Jurkat cell line is an immortalized line of human T-lymphocytes (Schneider et al., 1977), while the associated hybridoma cell line secretes antibodies against human CD4. CD4 is a membrane glycoprotein that is a co-receptor of T-cell receptors (TCRs) (Glatzová and Cebecauer, 2019), assisting in the communication with antigen-presenting cells. CD4 is also utilised by HIV-1 to gain entry into T-cells (Kwong et al., 1998).

Anti-CD4 antibodies have been used for *in vitro* T-cell and Jurkat cell activation, although this is typically done in conjunction with the OKT-3 antibody (Klammt et al., 2015).

5.2.1 Single-cell RNA-seq via Drop-seq
Jurkat cells with anti-CD4 hybridoma supernatant

We assessed the transcriptomic responses of untreated Jurkat cells and Jurkat cells treated with anti-CD4 hybridoma supernatant, which were subsequently processed via Drop-seq. No clustering by sample is observed for the differentially treated Jurkat cells at the single-cell level (Fig. 5.5). This suggests that the model system involving Jurkat and anti-CD4 hybridoma cells does not induce transcriptomic changes in Jurkat cells that are distinct enough to be detected at the single-cell level, demonstrating that this model system cannot be used for the development of our technology.

5.3 K-562 cells with OKT-9 hybridoma cells

The K-562 cell line is a human immortalised myelogenous leukemia cell line (Lozzio and Lozzio, 1975), while OKT-9 mouse hybridoma cells secrete antibodies that bind to the transferrin receptor (CD71) present on K-562 cells (Sutherland et al., 1981). CD71 plays a physiological role in cellular iron intake, is necessary for erythrocyte development and is implicated in carcinogenesis of various types of cancers (Chan et al., 2014b; Daniels et al., 2012), and thus we expected a detectable transcriptomic response upon binding of OKT-9 antibodies to K-562 cells.

Additionally, this model system has been previously used in droplet microfluidic assays utilising dual-colour normalised fluorescence readouts to detect antibody binding in a high-throughput manner (Shembekar et al., 2018).

5.4 Jurkat cells with OKT-3 hybridoma cells

The Jurkat cell line is an immortalized line of human T-lymphocytes (Schneider et al., 1977), while OKT-3 mouse hybridoma cells secrete antibodies that bind CD3,

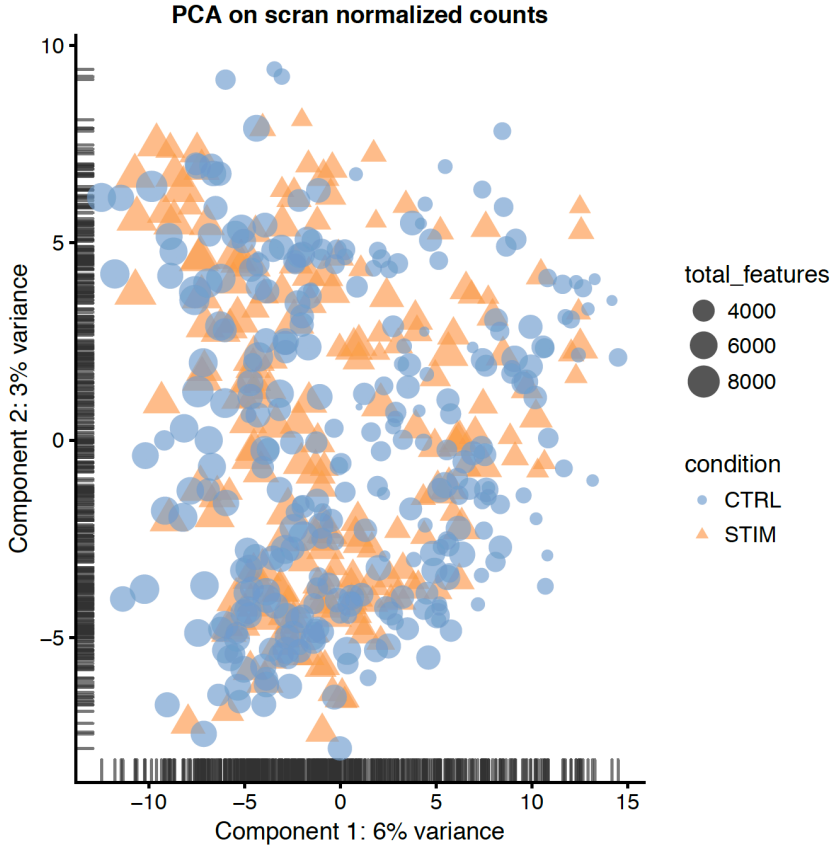


Figure 5.5: Dimensional reduction of gene expression data of Jurkat cells. Gene expression data of untreated Jurkat cells (cultured in media only) (blue circles) and Jurkat cells treated with anti-CD4 hybridoma supernatant (orange triangles), obtained by Drop-seq. The Principal Component Analysis was carried out on the scran normalised counts, with the first two principal components plotted. The size of the shapes indicate the number of total features, namely the number of genes that are detected per droplet. Data reproduced with the permission of Charles Girardot, EMBL Heidelberg.

a T-cell co-receptor that plays a role in T-cell activation. OKT-3 can activate T-cells (Norman, 1995), with this activation having been shown to be associated with a transcriptomic response detectable in bulk (Smeets et al., 2012).

5.5 Bulk RNA-seq

K-562 cells with OKT-9 hybridoma supernatant and Jurkat cells with OKT-3 hybridoma supernatant

In order to assess the suitability of the model systems involving K-562 cells with OKT-9 hybridoma cells and Jurkat cells with OKT-3 hybridoma cells, we conducted a bulk RNA-seq experiment. We sequenced 3 replicates each of K-562 cells treated with OKT-9 or H25B10 (negative control) hybridoma supernatants, and Jurkat cells treated with OKT-3 or H25B10 (negative control) hybridoma supernatants. Our aim was to test if there were transcriptomic changes present that were detectable in bulk.

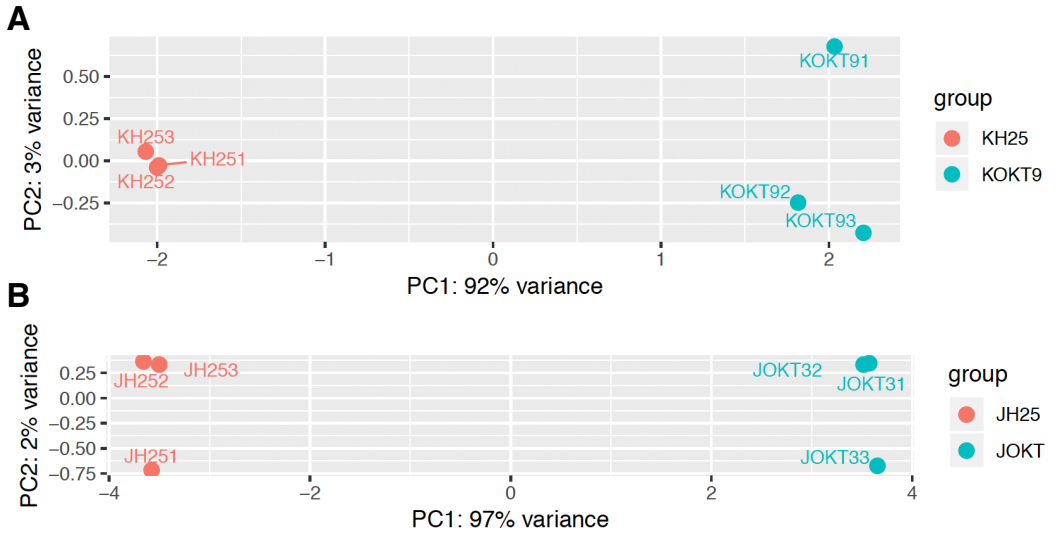


Figure 5.6: Bulk RNA-seq of two promising model systems. Gene expression data from bulk RNA-seq experiments were analysed via Principal Component Analysis (PCA) and the first two principal components were plotted. **A.** Clustering of gene expression data from K-562 cells treated with H25B10 hybridoma supernatant (red, "KH25") or with OKT-9 hybridoma supernatant (turquoise, "KOKT9"). **B.** Clustering of gene expression data from Jurkat cells treated with H25B10 hybridoma supernatant (red, "JH25") or with OKT-3 hybridoma supernatant (turquoise, "JOKT"). Data reproduced with the permission of Charles Girardot, EMBL Heidelberg.

For both model systems, we observed strong clustering of replicates based on treatment (Fig. 5.6). In particular, differentially-treated samples cluster at opposite ends of the x-axis, which represents the first principal component (PC1) and accounts for 92% and 97% of the variance for the K-562 and Jurkat systems respectively. This reflects that most of the variance present is due to the treatment with the different hybridoma supernatants, which illustrates that treatment of K-562 cells with OKT-9 hybridoma supernatant, and Jurkat cells with OKT-3 hybridoma supernatant does induce transcriptomic changes in the respective target cells which can be detected in bulk RNA-seq, laying the groundwork for single-cell transcriptomic analysis (Section 5.7).

5.6 Synthetic model systems

Two synthetic model systems were utilised, namely MCF-7 and A375-P cells treated with niclosamide, or with DMF (as a negative control).

The MCF-7 cell line is a human immortalised breast cancer cell line (Soule et al., 1973) which has been widely used in the study of breast cancer, while the A375-P cell line is a human malignant melanoma cell line (Welch et al., 1991).

Niclosamide is a small molecule drug that is used to treat parasitic infections, but as it modulates Wnt/β-catenin, mTORC1, STAT3, NF-κB and Notch signaling pathways, it has been repurposed for the treatment of other diseases, including

cancer (Barbosa et al., 2019; Chen et al., 2018). The interactions of both these cell lines with niclosamide have been previously characterised in the Connectivity Map, where niclosamide had a high average transcriptional impact score of 0.65, which is in the 94th percentile of all tested drugs present in the Connectivity Map, with the specific Transcriptional Activity Scores above 0.5 for both MCF-7 and A375 (Subramanian et al., 2017). This reflects that a large transcriptomic change was induced in both these cell types upon treatment with niclosamide.

The Connectivity Map also enabled us to identify a subset of genes that have been shown to be up- or down-regulated in MCF-7 and A375 cells upon niclosamide treatment, specifically by identifying genes with absolute Z-scores larger than 5 (Subramanian et al., 2017). Based on the Connectivity Map data, *TSC22D3* and *DDIT4* were up-regulated in MCF-7 cells upon niclosamide treatment, while *MAT2A* was down-regulated. Similarly, *INSIG1* and *ARHGEF2* were up-regulated in A375 cells treated with niclosamide.

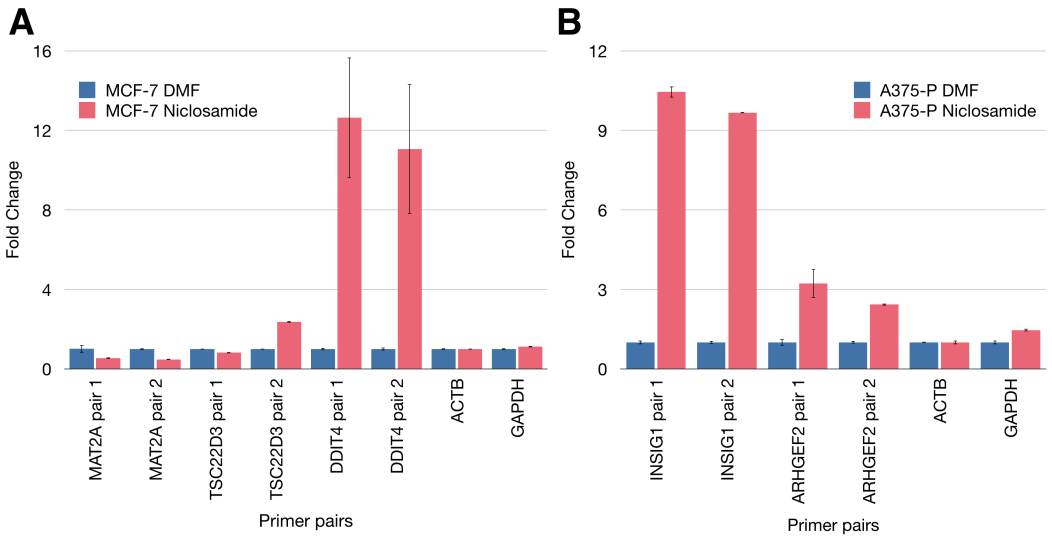


Figure 5.7: Relative gene expression in MCF-7 and A375-P cells upon treatment with niclosamide. Normalised relative expression of the genes of interest, *ACTB* and *GAPDH* in MCF-7 and A375-P cells upon treatment with niclosamide. Gene expression was determined via reverse transcription followed by quantitative real-time PCR (RT-qPCR), with the expression of each gene of interest assessed using two independent primer pairs. Gene expression was normalised to that of cells treated with dimethylformamide (DMF) (negative control) (blue) and to the expression of *ACTB*, a housekeeping gene. *GAPDH*, a second housekeeping gene, was also analysed. Two replicates were carried out for each sample - primer pair combination (n = 2), with the error bars indicating the standard deviation present. **A.** Normalised relative expression of *MAT2A*, *TSC22D3* and *DDIT4* in MCF-7 cells upon treatment with niclosamide (pink). **B.** Normalised relative expression of *INSIG1* and *ARHGEF2* in A375-P cells upon treatment with niclosamide (pink).

We designed qPCR primers against these genes and confirmed that the expected transcriptomic effects were detectable in cDNA obtained from MCF-7 and A375-P cells treated with niclosamide (Fig. 5.7). Specifically, *MAT2A* expression was reduced by 50% and an approximate 10-fold increase in *DDIT4* expression was noted for

MCF-7 cells treated with niclosamide (Fig. 5.7A). Similarly, an increase in *INSIG1* expression of over eight-fold and an approximate two-fold increase increase in *ARHGEF2* expression were observed for A375-P cells treated with niclosamide (Fig. 5.7B).

These suggested that a larger transcriptomic effect might be present and might be detectable with single-cell transcriptomics, and thus, these systems have been included in our 10X experiments as positive controls.

5.7 Studying multiple model systems with cell hashing

5.7.1 Cell hashing

As the 10X Genomics Single Cell 3' Gene Expression platform performs the best out of the three emulsion-based single-cell transcriptomic technologies (Zhang et al., 2019), we have utilised it to test if the transcriptomic changes seen in the target cell lines upon stimulation are detectable via single-cell transcriptomics. To minimise batch effects caused by the independent processing of different samples, we have utilised cell hashing (Stoeckius et al., 2018).

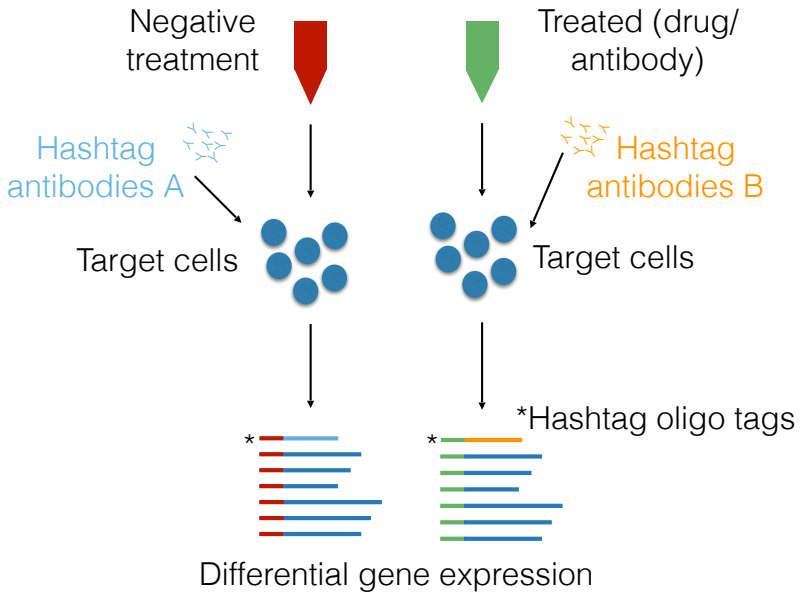


Figure 5.8: Cell hashing enables the identification of a desired model system. Two samples of target cells (blue) are differentially treated (red, green) before being tagged with different hashtag antibodies (light blue, orange). After sequencing, the sample identity of the cells can be elucidated, based on the presence of different oligo tags (as indicated by the asterisk). Based on the cell barcodes (red, green), the oligo tag sequences can be linked to the gene expression of individual single cells, to study differential gene expression.

Differentially-treated samples and samples from different model systems were all treated with antibodies against ubiquitously-expressed cell surface proteins, which have been linked to different oligonucleotide tags (Stoeckius et al., 2018). The

different samples were then pooled and processed via the 10X Genomics 3' Gene Expression v3 or v3.1 pipeline (Fig. 5.8). By sequencing both the cell transcriptomes and the cell hashtags, each specific transcriptome can be mapped back to a specific sample, and as all samples are processed in a single sequencing experiment, batch effects caused by processing are minimised. Cell hashing also permits the pooling and simultaneous analysis of multiple independent model systems, making it cost-effective.

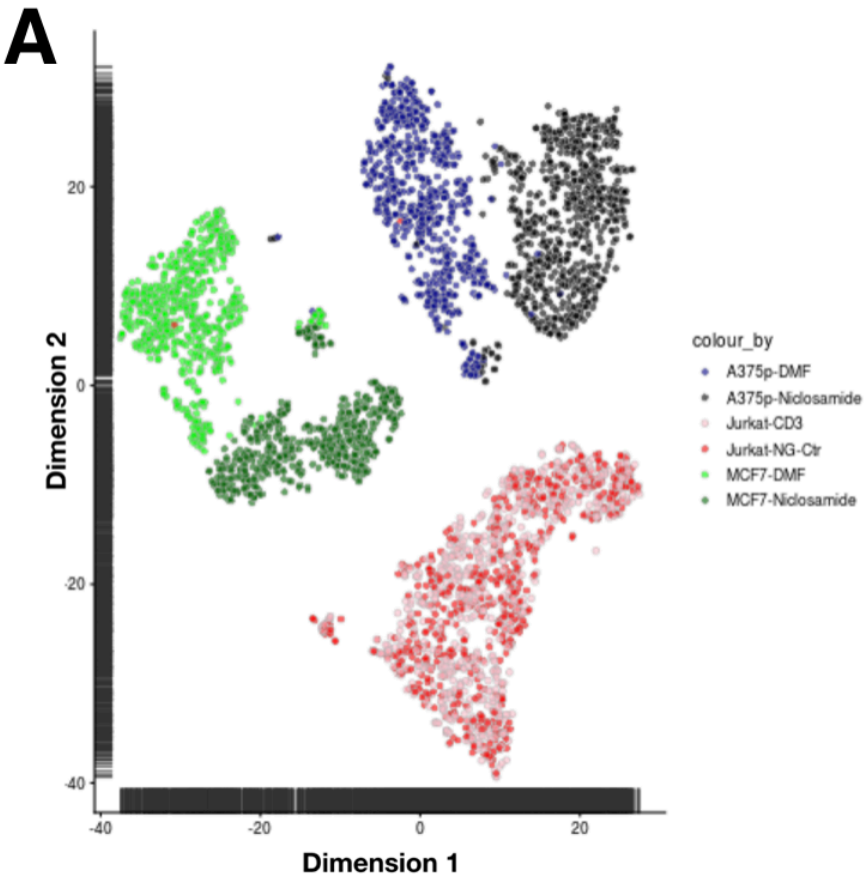
5.7.2 Single-cell RNA-seq with 10X Genomics

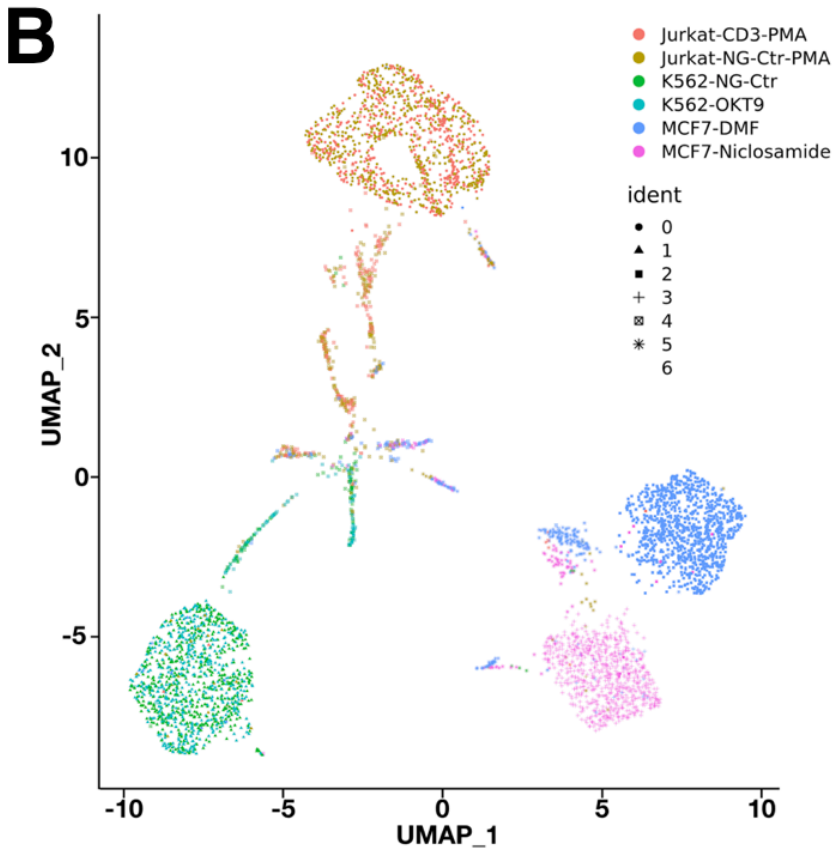
Over two independent experiments, we have studied six pairs of samples in total. We have included two positive model systems, which involve two cell lines, MCF-7 and A375-P, which were both treated with DMF (negative control) or niclosamide. We carried out two independent experiments with MCF-7 cells treated with niclosamide, where we compared the data obtained when the experiments were conducted with and without cell cycle synchronisation (Appendix B). While the systems do not involve cell-cell interactions, but rather the effect of a drug on cell transcriptomes, we believed that the strong transcriptional changes induced by niclosamide could be used as a control to study the number of cells needed to detect and identify a group of perturbed cells via single-cell transcriptomics (Chapter 6).

In addition, we had three pairs of samples from our model systems of interest, namely K-562 cells treated with anti-BSA antibody (negative control) or OKT-9 supernatant, as well as Jurkat cells treated with anti-His antibody (negative control) or anti-CD3 antibody, in the absence and presence of 10 ng/mL phorbol 12-myristate 13-acetate (PMA). PMA is used to stimulate T-cell activation, proliferation, and cytokine production (Weissman et al., 1986) and had previously been shown to increase the differential gene expression between untreated Jurkat cells and cells treated with anti-CD3 antibody (Smeets et al., 2012).

As the different model systems utilise human cell lines from different tissue types, we expect them to cluster separately upon dimensionality reduction and visualisation, due to cell line specific differences in gene expression. We observe this in the t-SNE and UMAP plots generated for both the first and second 10X experiments (Fig. 5.9), indicating that cell hashing does indeed permit distinguishing between differently treated and differently labelled samples. It also indicates that the single-cell transcriptomic analysis permits delimitation of cells from different cell lines.

MCF-7 cells treated with niclosamide cluster separately from MCF-7 cells treated with DMF, in the presence and absence of cell cycle synchronisation (Fig. 5.10A, B). We did not observe stark differences between the synchronised and unsynchronised samples, despite the fact that synchronisation was highly efficient, with unsynchronised samples having 69.0% of MCF-7 cells in G1, which increased



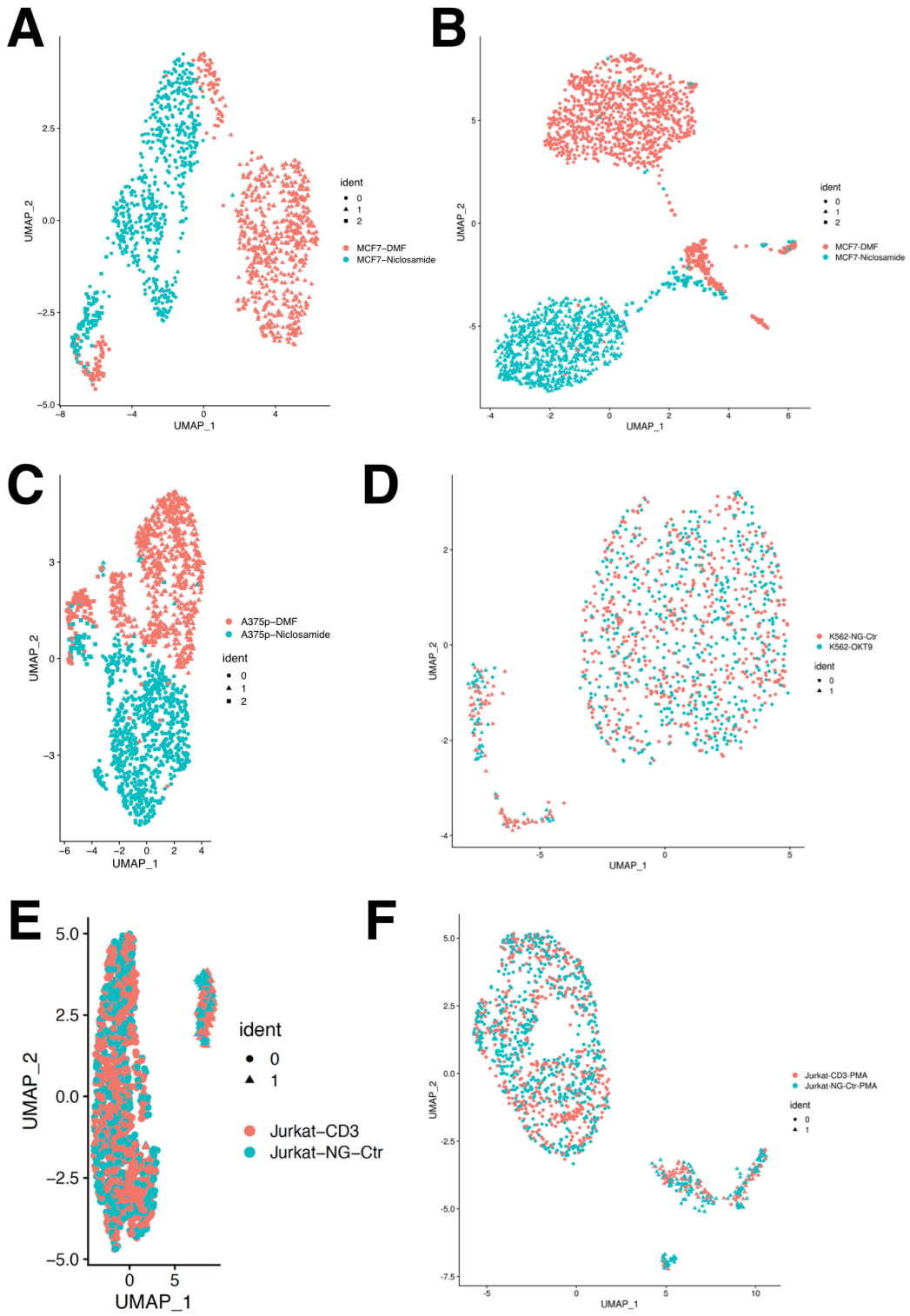


(see next page)

Figure 5.9: Distinguishing different cell types and treatments with cell hashing. Normalised gene expression data of different cell types treated with stimulants. **A.** t-distributed stochastic neighbourhood embedding (t-SNE) clustering of six samples from the first 10X and cell hashing experiment, comprising of A375-P cells treated with dimethylformamide (DMF, negative, blue) or with niclosamide (black), Jurkat cells treated with anti-His-antibody (negative, dark red) or with anti-CD3 antibody (pink), and unsynchronised MCF-7 cells treated with DMF (negative, light green) or with niclosamide (dark green). A perplexity of 50 was used. **B.** Six samples from the second 10X and cell hashing experiment visualised using Uniform Manifold Approximation and Projection (UMAP), comprising of Jurkat cells treated with anti-His-antibody (negative, mustard) or with anti-CD3 antibody (orange) in the presence of PMA, K-562 cells treated with anti-BSA antibody (negative, teal) or with OKT-9 supernatant (green), and synchronised MCF-7 cells treated with DMF (negative, blue) or with niclosamide (purple). The data was generated via the 10X Genomics 3' Gene Expression v3 pipeline. Data reproduced with the permission of Charles Girardot, EMBL Heidelberg.

to 95.9% upon synchronisation (Appendix B). Similarly, A375-P cells treated with niclosamide cluster separately from A375-P cells treated with DMF (Fig. 5.10C). These signify that niclosamide is able to induce a transcriptomic change in MCF-7 and A375-P cells that can be detected via single-cell transcriptomics, and these results are used in Chapter 6 to further examine the bioinformatic identification of positive events.

In contrast, no clustering by sample is apparent for K-562 cells treated with OKT-9 hybridoma supernatant (Fig. 5.10D), or for Jurkat cells treated with anti-CD3 antibody in the absence and presence of PMA (Fig. 5.10E, F). These suggest that the transcriptomic changes induced by these stimuli are not strong enough to be detected by single-cell transcriptomics, and that these two model systems cannot be used to further develop our technology. As such, we continued to search for a suitable model system.



(see next page)

Figure 5.10: Dimensional reduction of gene expression data from model systems. Normalised gene expression data of different cell types treated with stimulants, visualised using Uniform Manifold Approximation and Projection (UMAP). The different colours indicate the sample identity based on the detected hashtags, while different shapes indicate different identities as determined by Seurat. **A.** UMAP of unsynchronised MCF-7 cells treated with DMF (red) or with niclosamide (turquoise) as part of the first 10X experiment. **B.** UMAP of cell-cycle synchronised MCF-7 cells treated with DMF (red) or with niclosamide (turquoise) as part of the second 10X experiment. **C.** UMAP of A375-P cells treated with DMF (red) or with niclosamide (turquoise) as part of the first 10X experiment. **D.** UMAP of K-562 cells treated with anti-BSA antibody (red) or with OKT-9 hybridoma supernatant (turquoise) as part of the second 10X experiment.**E.** UMAP of Jurkat cells treated with anti-His antibody (turquoise) or with anti-CD3 antibody (red) as part of the first 10X experiment. **F.** UMAP of Jurkat cells treated with PMA, and with anti-His antibody (turquoise) or with anti-CD3 antibody (red) as part of the second 10X experiment. Data reproduced with the permission of Charles Girardot, EMBL Heidelberg.

5.8 Differentiated U937 cells with 60bca hybridoma cells

U937 cells are a human macrophage cell line, which can be induced to differentiate and express CD14 upon the addition of 100 nM 1,25-dihydroxyvitamin D3 (Baek et al., 2009). CD14 is a membrane co-receptor for the binding and detection of bacterial lipopolysccharides (LPS), and thus plays a crucial role in regulating cellular and subsequent immune responses to LPS (Zanoni et al., 2011). The 60bca mouse hybridoma cell line secretes antibodies that bind to CD14 (Wright et al., 1990), inhibiting the binding of the LPS-LBP protein complexes to CD14, preventing the release of tumour necrosis factor alpha (TNFα) (Wright et al., 1990). Thus, we expected the activation of differentiated U937 cells with LPS and LBP in the presence and absence of 60bca antibodies to result in differential transcriptomic responses.

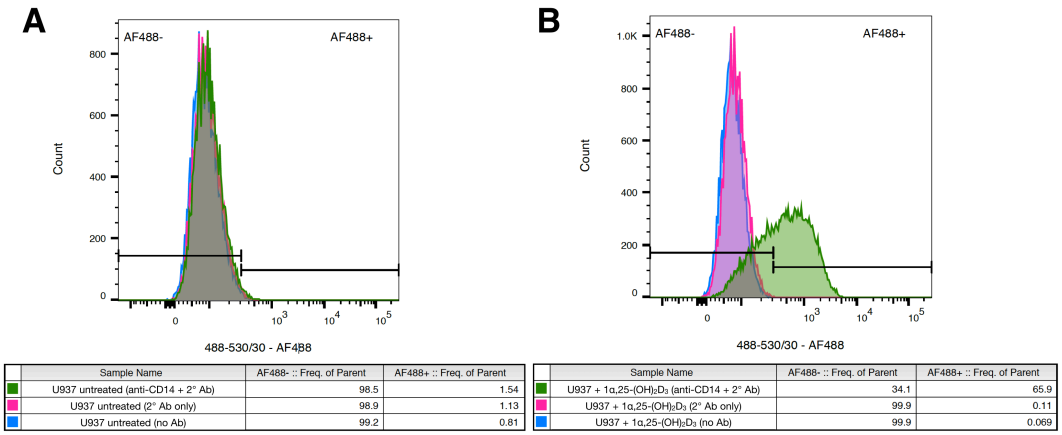
5.8.1 Differentiation of U937 cells

We have established the induction of CD14 expression in U937 cells. CD14 expression was detected in U937 cells treated with 100 nM 1,25-dihydroxyvitamin D3 for 24 hours (Fig. 5.11B, green), but not in untreated U937 cells (Fig. 5.11A, green). This fluorescence signal was absent in the absence of the primary antibody (mouse anti-human CD14 antibody) (Fig. 5.11B, blue and pink), demonstrating that it is indeed specific to CD14 expression.

5.8.2 Bulk mRNA-seq

U937 cells, LPS and 60bca hybridoma supernatant

In order to assess the suitability of U937 cells induced to express CD14 with 60bca hybridoma cells as a model system, we conducted a bulk RNA-seq experiment. We sequenced U937 cells induced to express CD14, in the presence of 500 ng/mL lipopolysccharides (LPS) and 1% human serum, which contains



Sample Name	AF488- :: Freq. of Parent	AF488+ :: Freq. of Parent
U937 untreated (anti-CD14 + 2° Ab)	98.5	1.54
U937 untreated (2° Ab only)	98.9	1.13
U937 untreated (no Ab)	99.2	0.81

Sample Name	AF488- :: Freq. of Parent	AF488+ :: Freq. of Parent
U937 + 1α,25-$(OH)_2D_3$ (anti-CD14 + 2° Ab)	34.1	65.9
U937 + 1α,25-$(OH)_2D_3$ (2° Ab only)	99.9	0.11
U937 + 1α,25-$(OH)_2D_3$ (no Ab)	99.9	0.069

Figure 5.11: Increased CD14 expression in U937 cells treated with 1,25-dihydroxyvitamin D3. A. Untreated U937 cells were incubated with no antibody (blue), with only 1000x diluted Alexa Fluor 488 goat anti-mouse IgG antibody (pink), or with 250x diluted mouse anti-human CD14 antibody and 1000x diluted Alexa Fluor 488 goat anti-mouse IgG antibody before flow cytometry analysis. **B.** U937 cells treated with 100 nM 1,25-dihydroxyvitamin D3 for 24 hours were incubated with no antibody (blue), with only 1000x diluted Alexa Fluor 488 goat anti-mouse IgG antibody (pink), or with 250x diluted mouse anti-human CD14 antibody and 1000x diluted Alexa Fluor 488 goat anti-mouse IgG antibody before flow cytometry analysis. Alexa Fluor 488 was detected at 488-530/30, x-axes.

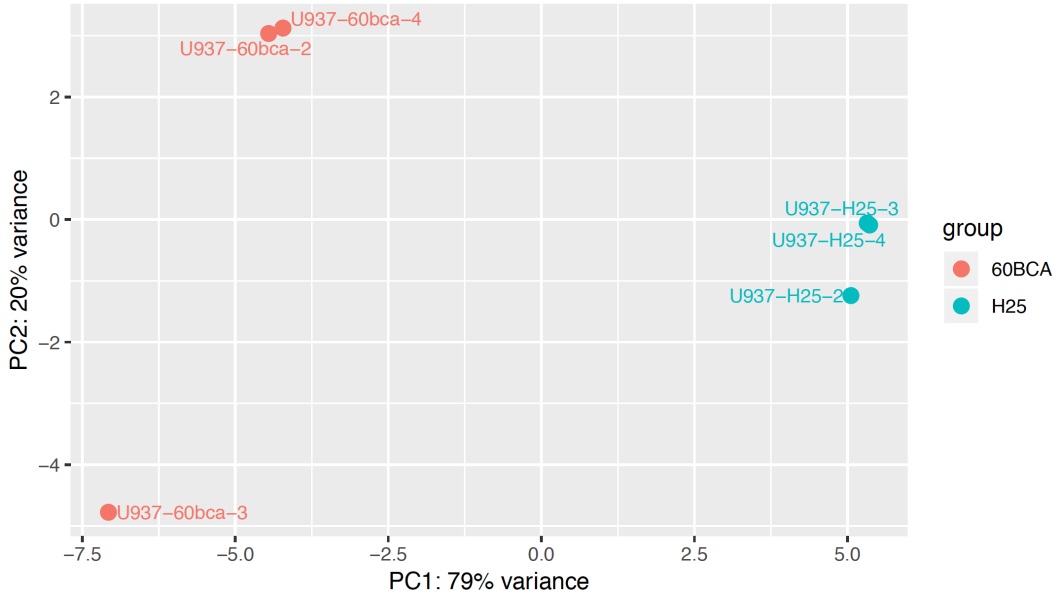


Figure 5.12: Bulk RNA-seq of CD14-expressing U937 cells in the presence of LPS and human serum. Gene expression data from a bulk RNA-seq experiment was processed with RUVSeq and analysed via Principal Component Analysis (PCA) and the first two principal components were plotted. Clustering of gene expression from U937 cells treated with 60bca hybridoma supernatant (red, "60BCA") or with H25B10 hybridoma supernatant (turquoise, "H25"), in the presence of 500 ng/mL LPS and 1% human serum. Data reproduced with the permission of Charles Girardot, EMBL Heidelberg.

lipopolysaccharide-binding protein (LBP). These cells were simultaneously treated with 60bca or H25B10 (negative control) hybridoma supernatants, to detect if there

is a detectable transcriptomic difference between samples treated with 60bca or H25B10 hybridoma supernatants.

We observed strong clustering of replicates based on treatment with different hybridoma supernatants (Fig. 5.12). In particular, differentially-treated samples cluster at opposite ends of the x-axis, which represents the first principal component (PC1) and accounts for 79% of the variance. This suggests that most of the variance present is due to the treatment of the U937 cells with the different hybridoma supernatants, in the presence of 500 ng/mL LPS and 1% human serum.

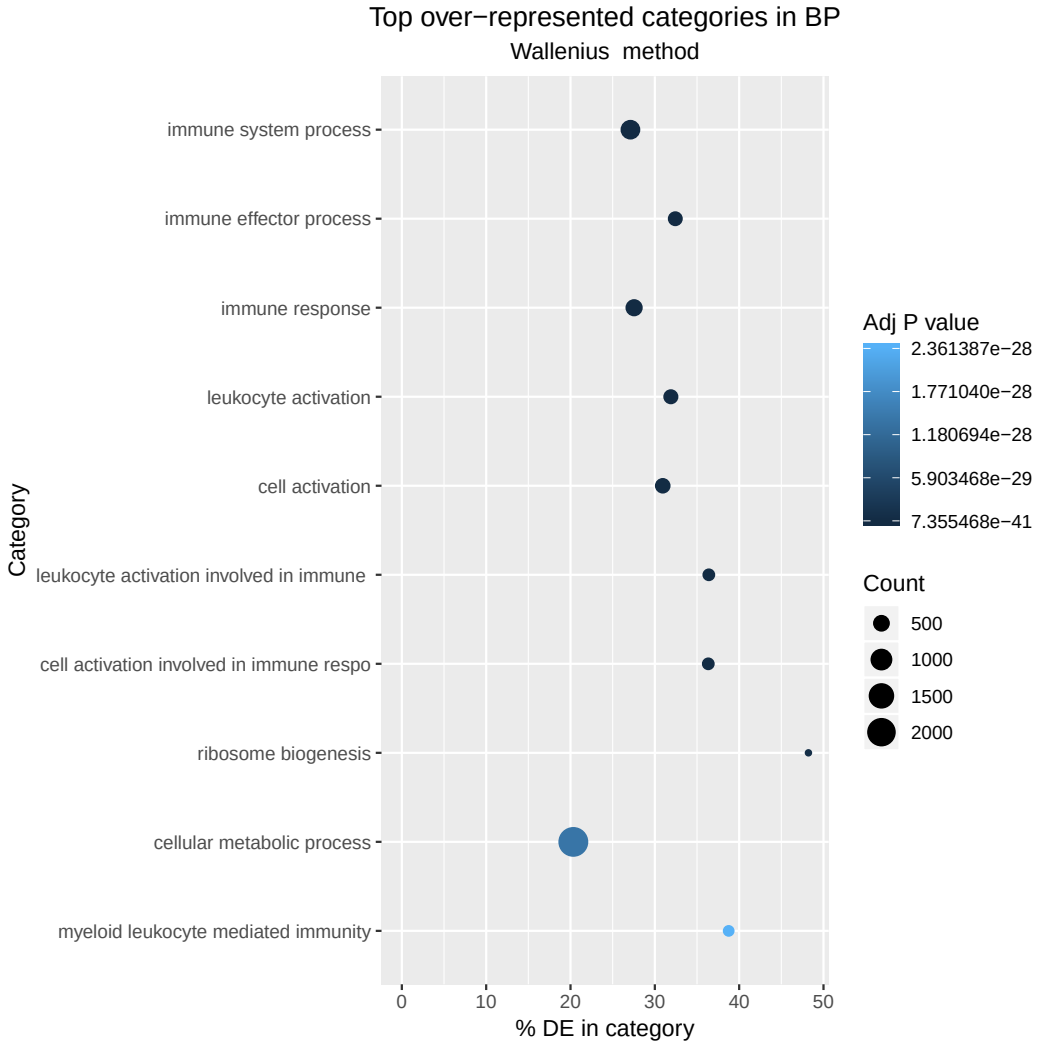


Figure 5.13: Gene ontology analysis of bulk RNA-seq data of CD14-expressing U937 cells in the presence of LPS and human serum. The top ten over-represented categories in Biological Processes (BP) are shown on the y-axis. The different shades of blue reflect the adjusted P-values, which were calculated using the Wallenius approximation method, while the size of the circles indicate the number of enriched genes that fall within each category. The x-axis reflects the percentage of genes within each category that were differentially expressed between the samples treated with 60bca and H25B10 hybridoma supernatants, both in the presence of 500 ng/mL LPS and 1% human serum. Data reproduced with the permission of Charles Girardot, EMBL Heidelberg.

Gene Ontology analysis was carried out on differentially expressed genes (FDR < 1%) (Appendix C, Table C.1) and the top ten over-represented categories include various cell activation and immune-associated processes (Fig. 5.13), which is biologically coherent with the LPS activation of U937 cells via CD14. U937 cells are human macrophage cells, and thus are part of the innate immune system, while CD14 plays a crucial role in regulating cellular and subsequent immune responses to LPS (Zanoni et al., 2011). Having these categories as over-represented amongst the genes differentially expressed between CD14-expressing U937 cells treated with 60bca or H25B10 hybridoma supernatants is consistent with 60bca antibodies having an effect on the binding of LPS-LBP complexes to CD14 and thus on subsequent CD14 signalling, suggesting that the differential transcriptomic signals detected are indeed the biological signals that we are interested in.

These suggest that the antibodies present in the 60bca hybridoma supernatant are capable of inhibiting LPS activation of CD14-expressing U937 cells, with transcriptomic effects detectable in bulk RNA-seq, laying the groundwork for future single-cell transcriptomic analyses (Section 5.9).

5.9 Studying the second batch of model systems with cell hashing

We have conducted a 10X Genomics Single Cell 3' Gene Expression experiment utilising cell hashing, where we studied two different model systems. We included six samples from the model system involving CD14-expressing U937 cells, LPS and 60bca or H25B10 hybridoma supernatants (Table 5.1).

Sample number	Sample name	LPS	Hybridoma supernatant
1	Untreated	No	Media only
2	H25B10 only	No	H25B10 supernatant
3	60bca only	No	60bca supernatant
4	LPS only	Yes, 500 ng/mL	Media only
5	LPS + H25B10	Yes, 500 ng/mL	H25B10 supernatant
6	LPS + 60bca	Yes, 500 ng/mL	60bca supernatant

Table 5.1: U937 samples included in the cell hashing experiment.

We have also included two samples from involving SK-BR-3 cells and Precision Antibody hybridoma cells (as outlined in Chapter 3), namely SK-BR-3 cells treated with H25B10 (negative control) or Precision Antibody hybridoma supernatant.

As the different model systems utilise human cell lines from different tissue types, we expect them to cluster separately upon dimensionality reduction and visualisation, due to cell line specific differences in gene expression. We observe this

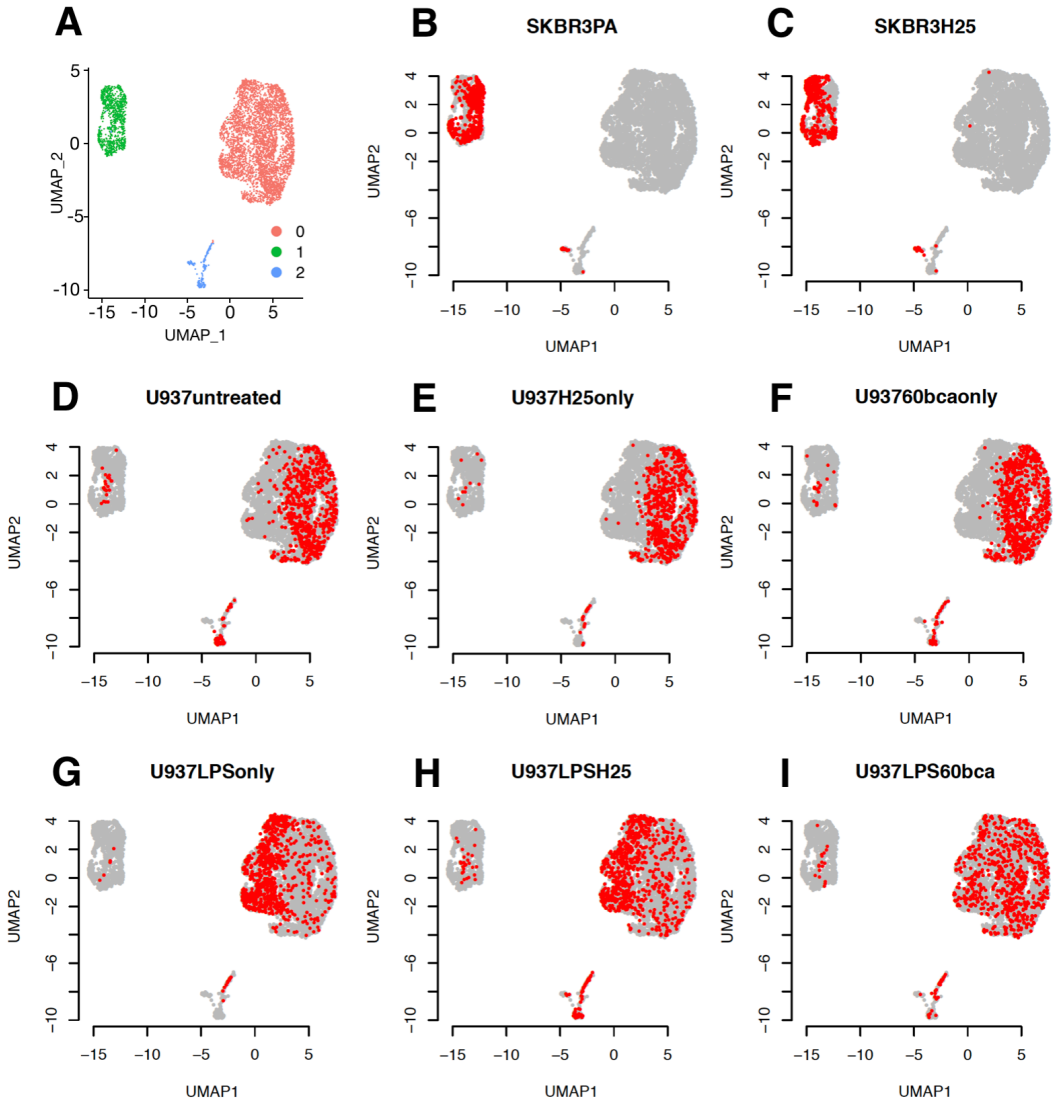


Figure 5.14: Distinguishing U937 and SK-BR-3 cells and different treatments with cell hashing. Gene expression data of different cell types treated with stimulants, visualised using Uniform Manifold Approximation and Projection (UMAP) with dimensions = 50 for dimensionality reduction, and resolution = 0.1 during clustering. **A.** Unsupervised clustering based on whole transcriptome data, of eight samples from the cell hashing experiment, comprising of SK-BR-3 cells treated with H25B10 supernatant or with supernatant from the Precision Antibody hybridomas and U937 cells triggered to express CD14, and treated as outlined in Table 5.1. Three different clusters were identified (red, green, blue). **B.-I.** Cells are differently highlighted based on their hashtag identity, with cells of the population of interest highlighted in red, and cells from the remaining samples shown in grey. SK-BR-3 cells treated with Precision Antibody hybridoma supernatant **(B)** and SK-BR-3 cells treated with H25B10 supernatant **(C)** were visualised. Untreated U937 cells **(D)**, U937 cells treated with only H25B10 hybridoma supernatant **(E)**, U937 cells treated with only 60bca hybridoma supernatant **(F)**, U937 cells treated with only LPS **(G)**, U937 cells treated with H25B10 hybridoma supernatant and LPS **(H)** and U937 cells treated with 60bca hybridoma supernatant and LPS **(I)** were visualised. The data was generated via the 10X Genomics 3' Gene Expression v3.1 pipeline. Data reproduced with the permission of Xiaoli Ma, EPFL.

in the generated UMAP plot, where SK-BR-3 cells and U937 cells cluster in distinct clusters, with SK-BR-3 cells largely clustering in Cluster 1 (green) and U937 cells clustering in Cluster 0 (red) (Fig. 5.14). This indicates that the cell hashing and 10X sequencing worked well.

The U937 cells from the six samples form two clusters upon unsupervised clustering, shown in red ("Cluster 0") and turquoise ("Cluster 1") in Figure 5.15A. Cells from samples that were not treated with LPS, but were either untreated (Fig. 5.15B), treated with H25B10 hybridoma supernatant (Fig. 5.15C), or with 60bca hybridoma supernatant (Fig. 5.15D), are largely localised in Cluster 0. In contrast, cells from samples that were treated with LPS only (Fig. 5.15E) or with LPS and H25B10 hybridoma supernatant (Fig. 5.15F) are largely localised in Cluster 1. Lastly, cells treated with both LPS and 60bca hybridoma supernatant appear to be spread evenly across both clusters (Fig. 5.15G).

These results suggest that the clustering observed is likely to be associated with LPS stimulation, with Cluster 0 (red) containing unstimulated cells and Cluster 1 (turquoise) containing cells stimulated by LPS (Fig. 5.15A). It should be noted that while the cells in samples treated with LPS only (Fig. 5.15E) or with LPS and H25B10 hybridoma supernatant (Fig. 5.15F) are largely localised in Cluster 1, a small number of cells are present in Cluster 0. This could be explained by the fact that only 66% of U937 cells express CD14 upon activation with 100 nM 1,25-dihydroxyvitamin D3 (Fig. 5.11B), such that not all U937 cells could be activated by LPS via CD14. The experiment could thus be further optimised by the fluorescence-activated cell sorting (FACS) of CD14-expressing cells prior to incubation with LPS and antibody supernatant.

The results also reinforce the fact that H25B10 hybridoma supernatant appears to induce little to no effect on target cell transcriptomes, which is optimal for a negative control. This can be noted from the fact that the clustering in the untreated (Fig. 5.15B) and H25B10 only samples (Fig. 5.15C) are highly similar. In addition, the clustering in the samples treated with LPS only (Fig. 5.15E) and LPS and H25B10 supernatant (Fig. 5.15F) are also extremely alike. The fact that the clustering present in the sample treated with only 60bca supernatant (Fig. 5.15D) is highly similar to that in the untreated and H25B10 only samples (Fig. 5.15B, C) suggests that the 60bca hybridoma supernatant has no effect on target cell transcriptomes in the absence of LPS, which makes biological sense, given that the 60bca antibodies bind CD14 and inhibit LPS activation via CD14.

In contrast, the distribution of cells treated with both LPS and 60bca hybridoma supernatant across both clusters shows that the 60bca hybridoma supernatant does indeed cause some inhibition of LPS activation (Fig. 5.15G), but that this inhibition is not complete, such that we do not see a clustering pattern reflective of no LPS activation. This partial inhibition is not optimal and could be further

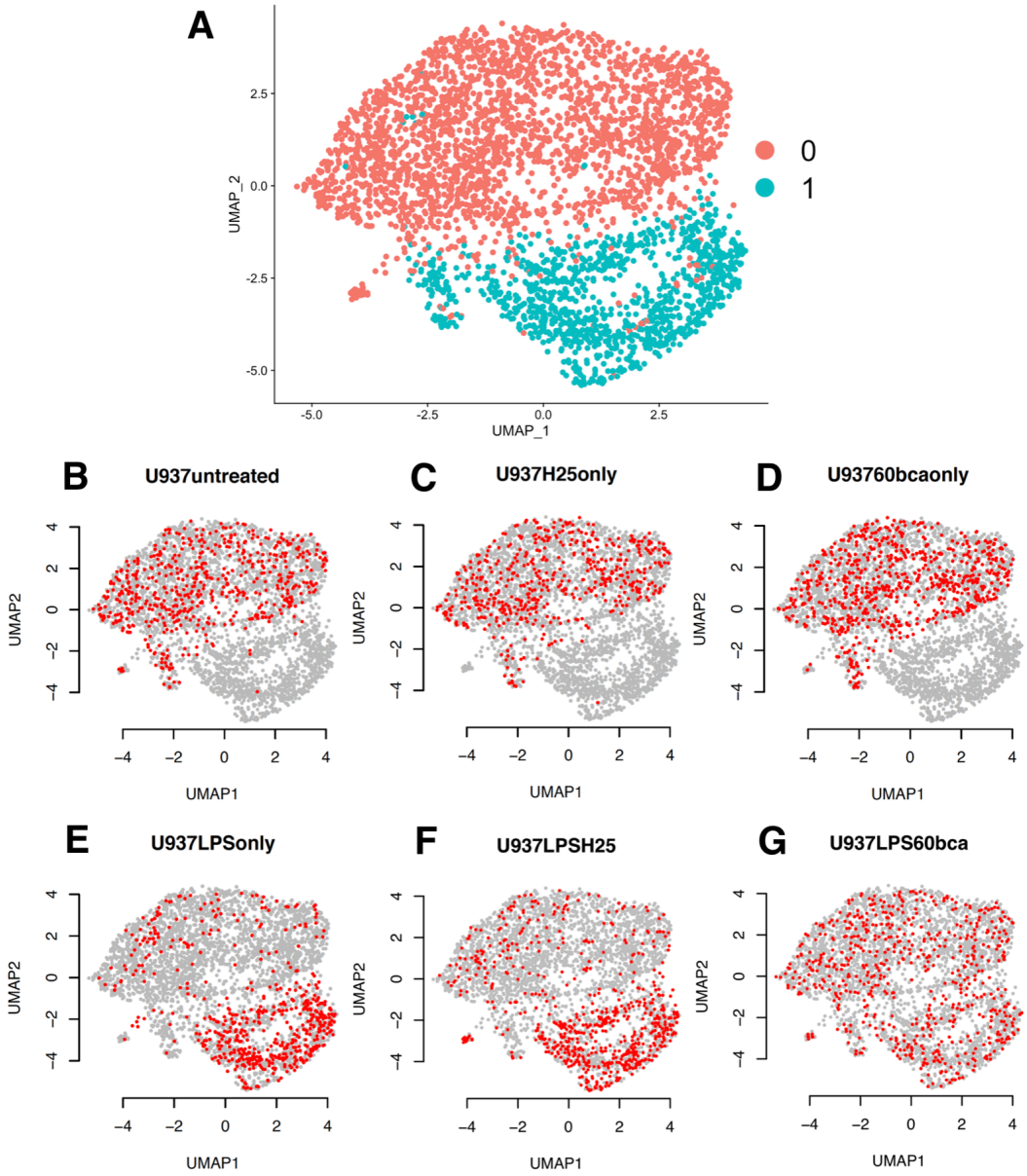


Figure 5.15: Dimensional reduction of gene expression data from U937 cells treated with LPS and H25B10 or 60bca hybridoma supernatant. Gene expression data of U937 cells treated with combinations of LPS, H25B10 hybridoma supernatant and 60bca hybridoma supernatant, as outlined in Table 5.1, and visualised using Uniform Manifold Approximation and Projection (UMAP) with dimensions = 50 for dimensionality reduction, and resolution = 0.1 during clustering. **A.** Unsupervised clustering based on whole transcriptome data of six samples involving U937 cells triggered to express CD14, and treated as outlined in Table 5.1. Two different clusters were identified (red, turquoise). **B.-G.** Cells are differently highlighted based on their hashtag identity, with cells of the population of interest highlighted in red, and cells from the remaining samples shown in grey. Untreated U937 cells **(B)**, U937 cells treated with only H25B10 hybridoma supernatant **(C)**, U937 cells treated with only 60bca hybridoma supernatant **(D)**, U937 cells treated with only LPS **(E)**, U937 cells treated with H25B10 hybridoma supernatant and LPS **(F)** and U937 cells treated with 60bca hybridoma supernatant and LPS **(G)** were visualised. Data reproduced with the permission of Xiaoli Ma, EPFL.

optimised by the modulation of LPS concentration, or by the adjustment of the antibody concentrations utilised for treatment. Nonetheless, this data suggests that treatment of CD14-expressing U937 cells with 60bca hybridoma supernatant (in the presence of LPS) does induce transcriptomic changes that are detectable via single-cell sequencing, paving the way for the use of this model system in subsequent experiments.

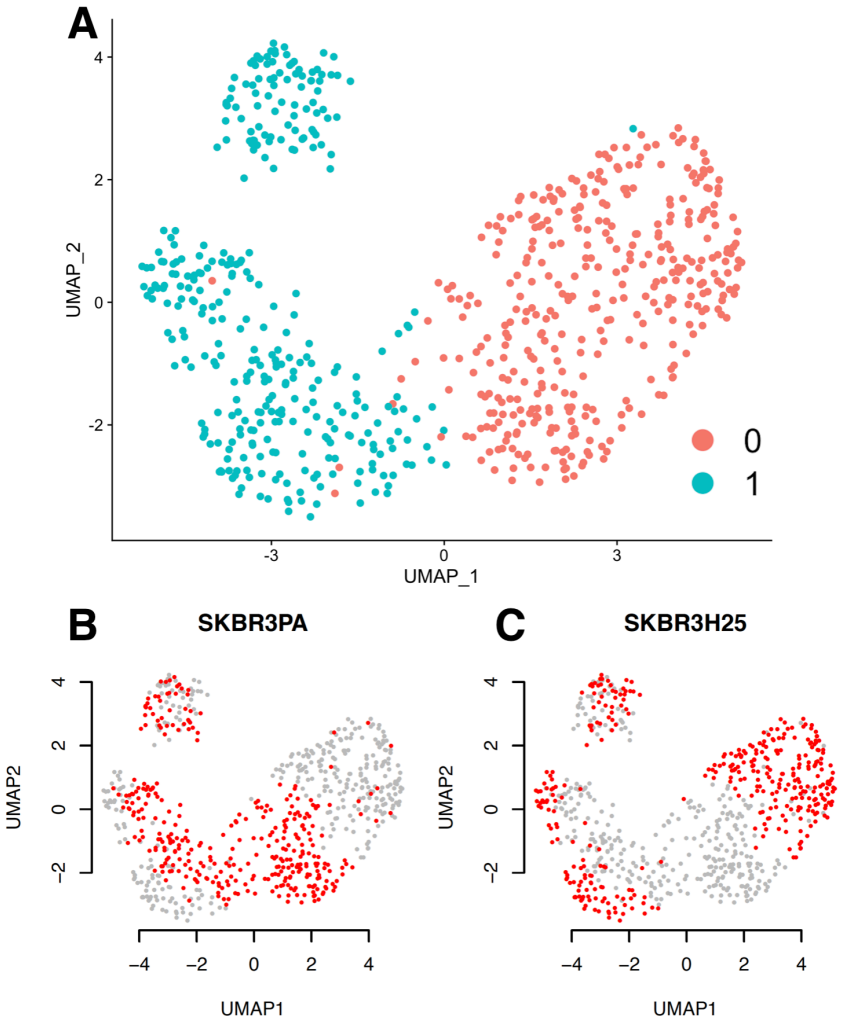


Figure 5.16: Dimensional reduction of gene expression data from SK-BR-3 cells treated with H25B10 and Precision Antibody hybridoma supernatant. Gene expression data of SK-BR-3 cells treated with H25B10 and Precision Antibody hybridoma supernatant, and visualised using Uniform Manifold Approximation and Projection (UMAP) with dimensions = 50 for dimensionality reduction, and resolution = 0.1 during clustering. **A.** Unsupervised clustering based on whole transcriptome data of two samples involving SK-BR-3 cells. Two different clusters were identified (red, turquoise). **B.** SK-BR-3 cells treated with Precision Antibody hybridoma supernatant are shown in red, while SK-BR-3 cells treated with H25B10 hybridoma supernatant are visualised in grey, based on the hashtag identity of the two samples. **C.** SK-BR-3 cells treated with H25B10 hybridoma supernatant are shown in red, while SK-BR-3 cells treated with Precision Antibody hybridoma supernatant are visualised in grey, based on the hashtag identity of the two samples. Data reproduced with the permission of Xiaoli Ma, EPFL.

SK-BR-3 cells treated with Precision Antibody or H25B10 hybridoma supernatant form two clusters upon unsupervised clustering, shown in red ("Cluster 0") and turquoise ("Cluster 1") in Figure 5.16A, but these clusters do not correlate with sample identity. However, some clustering based on sample identity did take place (Fig. 5.16B, C), suggesting that there are indeed transcriptomic differences between the two populations that can be detected by single-cell transcriptomic analyses. Therefore, we are confident that a subset of PA hybridoma cells secrete antibodies that bind to, and trigger transcriptomic changes in SK-BR-3 cells. To further study individual hybridoma cells and the effects of their antibodies on SK-BR-3 transcriptomes, we will co-encapsulate individual hybridoma cells together with SK-BR-3 target cells (as outlined in Chapter 4). This will permit the accumulation of antibodies from specific single hybridoma cells to functional concentrations within droplets in a matter of hours, and will permit the detection of transcriptomic effects triggered by antibodies secreted from a single hybridoma cell. Therefore, this experiment is planned for the future.

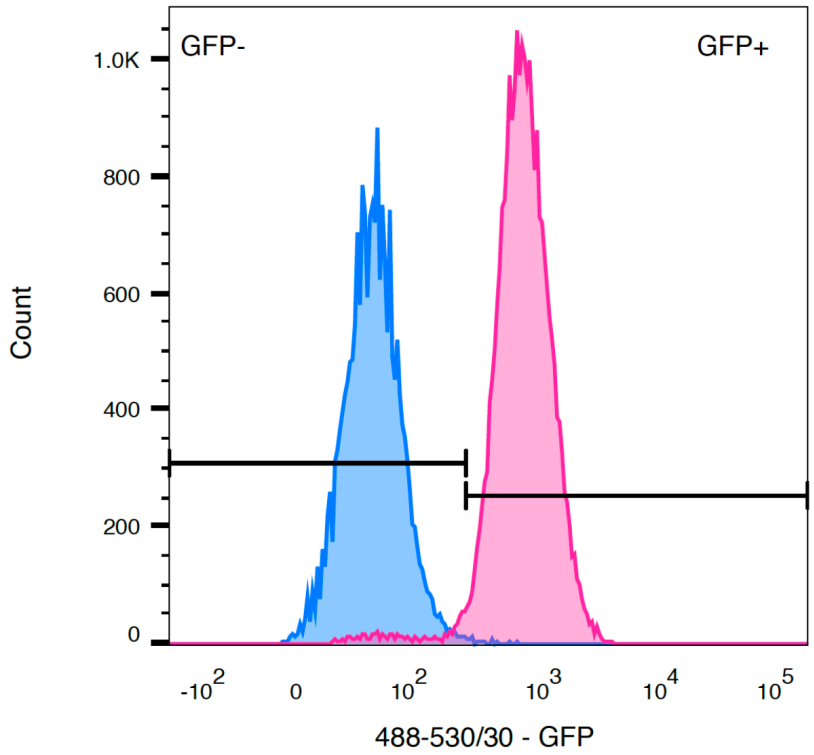
5.10 CHO-GLP1R-GFP cells with HEK293T-mCherry-Exendin-4 cells

While the focus of this chapter is primarily on cell-cell interactions involving antibodies, the technology we aim to develop would be equally applicable to cell-cell interactions involving direct cell contact. Therefore, we also tested a model system involving CHO-GLP1R-GFP cells with HEK293T-mCherry-Exendin-4 cells, which interact via direct cell contact.

The glucagon-like peptide 1 receptor (GLP1R) is a GPCR protein receptor involved in the regulation of insulin secretion, while Exendin-4 is an agonist of GLP1R (Thorens et al., 1993). The CHO-GLP1R-GFP cell line is a Chinese Hamster Ovary cell line that expresses human GLP1R, and expresses green fluorescent protein (GFP) under the control of a CRE promoter, such that GFP is expressed upon the activation of the human GLP1R (Zhang et al., 2015). The HEK293T-mCherry-Exendin-4 cell line is a human embryonic kidney cell line that expresses mCherry and Exendin-4 fused to the platelet-derived growth factor receptor (PDGFR) transmembrane domain, to anchor Exendin-4 to the cell surface (Zhang et al., 2015). The activation of CHO-GLP1R-GFP cells with Exendin-4 results in increased expression of GFP (Fig. 5.17), while the co-cultivation of HEK293T-mCherry-Exendin-4 cells with CHO-GLP1R-GFP induces GFP expression in the latter, which is not observed when CHO-GLP1R-GFP cells are co-cultivated with wild-type HEK293T cells (Fig. 5.18). While the GFP expression acts as a marker gene that can be easily identified, we expected further changes in gene expression in CHO-GLP1R-GFP cells to be triggered by the presence of Exendin-4.

5.10.1 Activation of CHO-GLP1R-GFP cells with 10 µM Exendin-4

The treatment of CHO-GLP1R-GFP cells with 10 µM Exendin-4 for 6 hours resulted in the increased expression of GFP, as detected via flow cytometry (Fig. 5.17, pink).

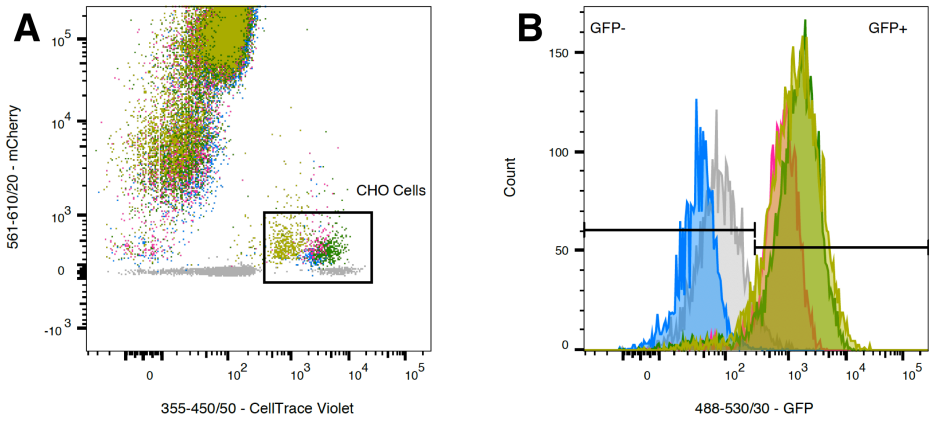


	Sample Name	GFP- :: Freq. of Parent	GFP+ :: Freq. of Parent
	CHO-GLP1R-GFP + 10 µM Exendin-4	3.35	96.6
	CHO-GLP1R-GFP negative	99.2	0.79

Figure 5.17: GFP expression in CHO-GLP1R-GFP cells treated with 10 µM Exendin-4. CHO-GLP1R-GFP cells were incubated for 6 hours either in regular media (negative control, blue) or with 10 µM Exendin-4 (pink), before analysis with flow cytometry. GFP expression was detected at 488-530/30, x-axis.

5.10.2 Co-cultivation of CHO-GLP1R-GFP cells with HEK293T-mCherry-Exendin-4 cells

Similarly, the co-cultivation of CHO-GLP1R-GFP cells with HEK293T-mCherry-Exendin-4 cells for over 6 hours resulted in the expression of GFP, as detected via flow cytometry (Fig. 5.18, 6 h: pink, 10 h: green and 24 h: mustard), while this GFP signal was not detected for CHO-GLP1R-GFP cells that were mixed with HEK293T-mCherry-Exendin-4 cells immediately prior to flow cytometry (Fig. 5.18, blue), or for CHO-GLP1R-GFP cells co-cultivated with HEK293T-WT cells (Fig. 5.18, grey). While the GFP expression is detectable after 6 hours, it should be noted that GFP expression appears to be even higher after 10 hours and after 24 hours (Fig. 5.18).



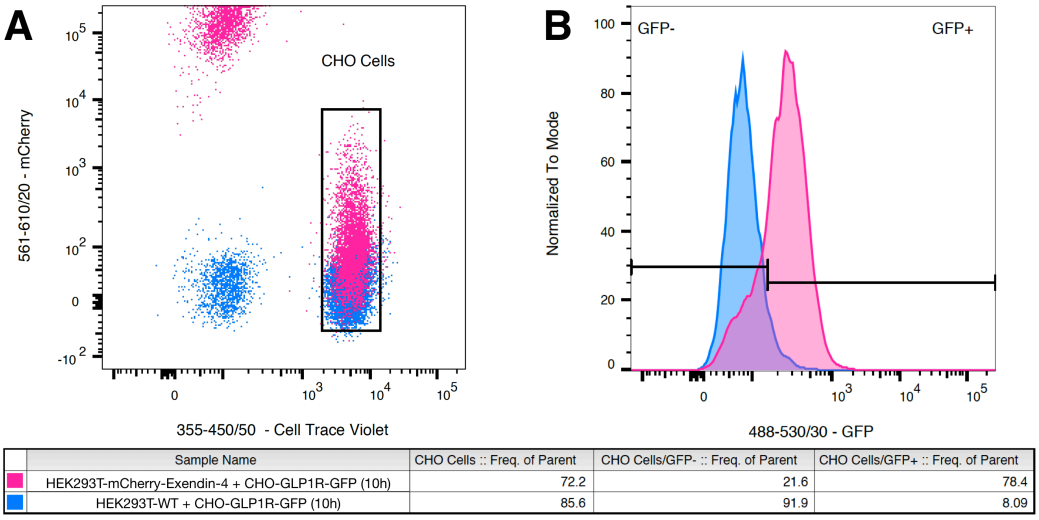
	Sample Name	CHO Cells :: Freq. of Parent	CHO Cells/GFP- :: Freq. of Parent	CHO Cells/GFP+ :: Freq. of Parent
	HEK293T-mCherry-Exendin-4 + CHO-GLP1R-GFP (24h)	4.86	6.75	93.2
	HEK293T-mCherry-Exendin-4 + CHO-GLP1R-GFP (10h)	4.18	6.53	93.5
	HEK293T-mCherry-Exendin-4 + CHO-GLP1R-GFP (6h)	2.46	7.84	92.2
	HEK293T-mCherry-Exendin-4 + CHO-GLP1R-GFP (0h)	2.22	99.4	0.55
	HEK293T-WT + CHO-GLP1R-GFP (30h)	3.39	92.6	7.42

Figure 5.18: GFP expression in CHO-GLP1R-GFP cells co-cultivated with HEK293T-WT or HEK293T-mCherry-Exendin-4 cells. CHO-GLP1R-GFP cells were stained with CellTrace™ Violet, seeded into plates containing HEK293T-WT cells and incubated for 30 hours (negative control, grey) or into plates with HEK293T-mCherry-Exendin-4 cells and incubated for 0 hours (blue), 6 hours (pink), 10 hours (green) or 24 hours (mustard). Cells were then harvested for flow cytometry. **A.** CHO-GLP1R-GFP cells were identified as the population of cells that were positive for CellTrace™ Violet and negative for mCherry (gate labelled CHO Cells). mCherry expression was detected at 561-610/20, y-axis, and CellTrace™ Violet was detected at 355-450/50, x-axis. **B.** The CHO cell populations were then examined for GFP expression, which was detected at 488-530/30, x-axis.

5.10.3 Co-encapsulation of CHO-GLP1R-GFP cells with HEK293T-mCherry-Exendin-4 cells

We have also carried out a similar assay in droplets, where we co-encapsulated CHO-GLP1R-GFP cells with HEK293T-mCherry-Exendin-4 cells or with WT HEK293T cells, at $\lambda = 1$ for each cell type. Cells were then analysed via flow cytometry 10 hours later. We observed that the CHO-GLP1R-GFP cells that had been co-encapsulated with HEK293T-mCherry-Exendin-4 cells (Fig. 5.19, pink) had a higher GFP expression level than those co-encapsulated with WT HEK293T cells (Fig. 5.19, blue). This demonstrates that the direct cell interaction can take place in droplets, and does result in the expected phenotypic change. Therefore, we expect that co-encapsulation and incubation in droplets will similarly lead to a detectable change in cell transcriptomes.

While it is not possible to quantitatively compare results across different flow cytometry experiments, it should be noted that the increase in GFP expression detected after 10 hours in the co-encapsulation experiment appears to be smaller than the increase seen after 10 hours of co-cultivation (Fig. 5.18, green and Fig. 5.19, pink). This is likely to be due in part to the Poisson statistics involved in the co-encapsulation of two cell types. While we expect each droplet to



Sample Name	CHO Cells :: Freq. of Parent	CHO Cells/GFP- :: Freq. of Parent	CHO Cells/GFP+ :: Freq. of Parent
HEK293T-mCherry-Exendin-4 + CHO-GLP1R-GFP (10h)	72.2	21.6	78.4
HEK293T-WT + CHO-GLP1R-GFP (10h)	85.6	91.9	8.09

Figure 5.19: GFP expression in CHO-GLP1R-GFP cells co-encapsulated with HEK293T-WT or HEK293T-mCherry-Exendin-4 cells. CHO-GLP1R-GFP cells were stained with CellTrace™ Violet and co-encapsulated with HEK293T-mCherry-Exendin-4 cells or with WT HEK293T cells. Cells were incubated in droplets for 10 hours, before the emulsion was broken and the cells were harvested for flow cytometry. **A.** CHO-GLP1R-GFP cells were identified as the population of cells that were positive for CellTrace™ Violet and negative for mCherry (gate labelled CHO Cells). mCherry expression was detected at 561-610/20, y-axis, and CellTrace™ Violet was detected at 355-450/50, x-axis. **B.** The CHO cell populations were then examined for GFP expression, which was detected at 488-530/30, x-axis. The scale of the y-axis has been normalised to the mode.

contain an average of one cell of each cell type, given a $\lambda = 1$ for each cell type, Poisson distribution predicts that 36.8% of droplets containing one or more CHO-GLP1R-GFP cells would not contain any HEK293T cells. As such, we expect a subset of CHO-GLP1R-GFP cells to not encounter any HEK293T-mCherry-Exendin-4 cells, and thus to not exhibit increased GFP expression. This would explain the non-symmetric curve present in Figure 5.19B.

However, the absence of a solid substrate in droplets could also cause changes in gene expression of adherent cells, such as CHO-GLP1R-GFP. This suggests that further optimisation of in-droplet experiments would be required, before utilising this model system for further co-encapsulation experiments. In addition, as the GFP expression increases with the time that CHO-GLP1R-GFP and HEK293T-mCherry-Exendin-4 cells are permitted to interact (Fig. 5.18), it would be wise to assess cell viability in droplets in conjunction with GFP expression, to maximise GFP signal while minimising cell death.

These experiments lay the foundation for future experiments with the CHO-GLP1R-GFP model system. The activation of CHO-GLP1R-GFP cells with Exendin-4 in plates permits the generation of bulk and single-cell RNA-seq data, while the activation of CHO-GLP1R-GFP cells via co-cultivation and

co-encapsulation with HEK293T-mCherry-Exendin-4 cells lay the groundwork for subsequent experiments combining co-encapsulation and Drop-seq.

5.10.4 Bulk mRNA-seq

CHO-GLP1R-GFP cells with Exendin-4

In order to assess the suitability of CHO-GLP1R-GFP cells with HEK293T-mCherry-Exendin-4 to be a model system, we conducted a bulk RNA-seq experiment, where untreated CHO-GLP1R-GFP cells and CHO-GLP1R-GFP cells treated with 10 µM Exendin-4 were sequenced in bulk.

We observed strong clustering of replicates based on treatment, both when GFP was excluded (Fig. 5.20A) and when it was included in the analysis (Fig. 5.20B). In particular, differentially-treated samples cluster at opposite ends of the x-axis, which represents the first principal component (PC1) and accounts for 99% of the variance. This suggests that most of the variance present is due to the treatment of CHO-GLP1R-GFP cells with Exendin-4, independently of GFP expression.

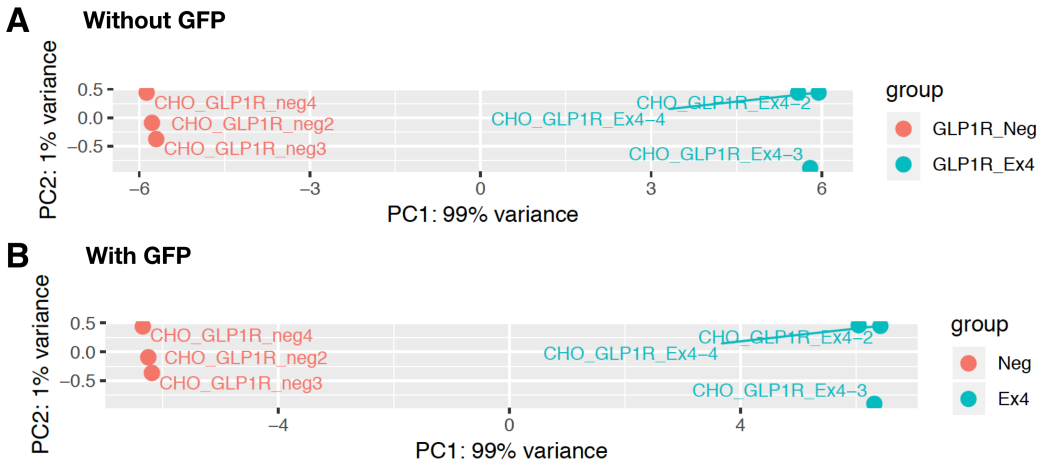


Figure 5.20: Bulk RNA-seq of CHO-GLP1R-GFP cells in the presence and absence of Exendin-4. **A.** Gene expression data, with the exclusion of GFP expression, were processed with RUVSeq and analysed via Principal Component Analysis (PCA) and the first two principal components were plotted. Clustering of gene expression data from CHO-GLP1R-GFP cells treated with Exendin-4 for 6 hours (turquoise, "GLP1R_Ex4") and untreated CHO-GLP1R-GFP cells (red, "GLP1R_Neg") are shown. **B.** Gene expression data, with GFP expression included, were processed with RUVSeq and analysed via Principal Component Analysis (PCA) and the first two principal components were plotted. Clustering of gene expression data from CHO-GLP1R-GFP cells treated with Exendin-4 for 6 hours (turquoise, "Ex4") and untreated CHO-GLP1R-GFP cells (red, "Neg") are shown. Data reproduced with the permission of Charles Girardot, EMBL Heidelberg.

In addition, GFP was also detected amongst the top differentially expressed genes when it was included in the analysis ("*TurboGFP*", Appendix C, Table C.2), which lends weight to the results of the bulk sequencing experiment. This suggests that GFP could be utilised as a positive control in subsequent sequencing experiments, where we expect to see higher GFP expression in CHO-GLP1R-GFP

cells in which GLP1R has been activated. The top differentially expressed genes from the bulk sequencing experiment could also be utilised for more effective identification of activated cells.

5.10.5 Single-cell RNA-seq via Drop-seq CHO-GLP1R-GFP cells with Exendin-4

We assessed the transcriptomic responses of untreated CHO-GLP1R-GFP cells and CHO-GLP1R-GFP treated for 24 hours with 10 µM Exendin-4, which were subsequently processed via Drop-seq. As the two populations cluster separately (Fig. 5.21), this suggests that the model system involving CHO-GLP1R-GFP and HEK293T-mCherry-Exendin-4 cells could be further utilised for the development of our technology.

Figure 5.21: Dimensional reduction of Drop-seq gene expression data of CHO-GLP1R-GFP cells. Gene expression data of untreated CHO-GLP1R-GFP cells (cultured in media only) ("CTRL", blue circles) and CHO-GLP1R-GFP cells treated with 10 µM Exendin-4 for 24 hours ("STIM", orange triangles). The Principal Component Analysis was carried out on the raw log counts after QC, with the first two principal components plotted. The size of the shapes indicate the number of total features, namely the number of genes that are detected per droplet. Data reproduced with the permission of Charles Girardot, EMBL Heidelberg.

The two populations separate along the second principal component (Fig. 5.21, y-axis), which accounts for 2% of the population variance. The first principal component corresponds to 64% of the variance and this appears to correlate to

the total features detected. This implies that the variance being picked up in the first principal component is technical rather than biological, suggesting that further normalisation may be required.

In addition, GFP was also detected amongst the top differentially expressed genes ("*TurboGFP*", Appendix C, Table C.3), supporting the idea that this could be used as a marker gene to identify activated CHO-GLP1R-GFP cells.

Despite the fact that different incubation periods with Exendin-4 were utilised for the bulk sequencing experiment and for this Drop-seq experiment, we detected three genes apart from GFP that were highly differentially expressed in both experiments (Appendix C, Tables C.2 and C.3), namely *Thbd*, *Tfap2a* and *Scarb1*.

These data suggests that the model system involving CHO-GLP1R-GFP and HEK293T-mCherry-Exendin-4 cells is a promising one, with the detection of differential gene expression induced by Exendin-4 at a single-cell level after cells were treated with Exendin-4 for 24 hours. Optimally, we would have similar data for 6 hours of Exendin-4 treatment, and potentially with the 10X Genomics 3' Gene Expression platform.

Nonetheless, we believe that the system could be further utilised in the development of our co-encapsulation technology. This would be particularly interesting in the context of cell-cell interactions where both the stimulant and the receptor are present on cell surfaces, such that cells must come into direct contact for the interaction to take place.

5.11 Conclusion

In this chapter, we have tested various potential model systems in a bid to find a suitable one in which to establish the technology.

The systems involving CHO-GLP1R-GFP and HEK293T-mCherry-Exendin-4 cells, and CD14-expressing U937 cells with 60bca hybridoma cells (in the presence of LPS) appear to induce transcriptomic changes in the target cells that are detectable by single-cell transcriptomic sequencing. This makes them well-suited for subsequent co-encapsulation experiments, which will enable the development and optimisation of a technology for high-throughput transcriptomic analyses of cell-cell interactions.

In addition, the GLP1R-Exendin-4 system requires direct cell contact for interaction, while the U937-60bca system involves antibody secretion and binding, such that the interaction is indirect. The fact that we have two model systems that involve different types of cell-cell interactions will enable us to ensure the suitability of our technology for both categories of cell-cell interactions.

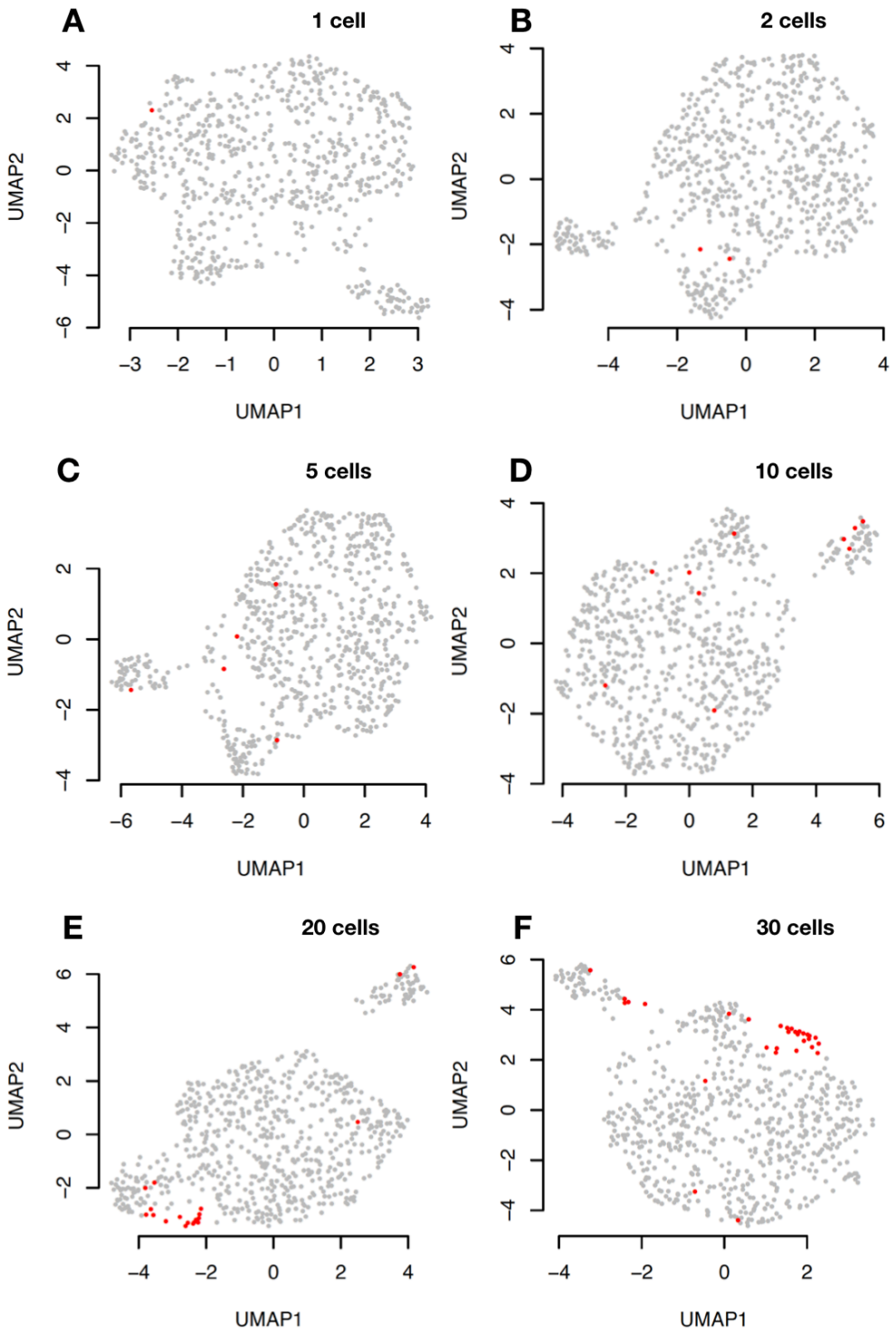
6 Bioinformatic identification of positive events

In order to study cell-cell interactions via single-cell transcriptomics, we would like to identify specific transcriptomic perturbations in the target cells and associate them back to specific stimulant cells. This raises a question about the minimum number of perturbed cells required to identify a perturbation against the background of unstimulated cells. Ideally, we hope to be able to identify single transcriptomes of interest, and thus detect single stimulant cells associated with these transcriptomes.

To further investigate the minimum number of cells required to detect a perturbation, we have carried out downsampling analyses. As we do not have an antibody model system in which treated and untreated cells cluster separately upon visualisation of single-cell transcriptomic data, we carried out these analyses on the model systems involving MCF-7 and A375-P cells that were treated with DMF (negative control) or with niclosamide (Fig. 5.10A, C). While the effect size of different model systems may vary, such that the conclusions from this analysis may not be completely applicable to all other model systems, we believe that this acts as a good first estimate.

For both cell types, all untreated cells were included, but a subset of the cells treated with niclosamide were spiked in into the single-cell dataset to be analysed. Subsequently, clustering was carried out via Seurat and the datasets were visualised with UMAP. The identity of the cells are known beforehand, such that the cells that have been spiked in can be identified after clustering. This was repeated for 1, 2, 5, 10, 20, 30, 50, 100, 200, 300, 400 and 500 cells treated with niclosamide, in the presence of 683 MCF-7 cells treated with DMF (negative control) (Fig. 6.1) or 763 A375-P cells treated with DMF (negative control) (Fig. 6.2).

As the number of treated cells included increases, the boundary between the untreated and treated cells becomes more defined, with separate clusters of treated cells becoming increasingly apparent (Figures 6.1 and 6.2). However, it must be noted that even between these two model systems, which involve two different cell types treated with the same drug, the number of cells at which one or more clusters appear differ. For MCF-7 cells treated with niclosamide, a separate cluster is


A
1 cell
B
2 cells
C
5 cells
D
10 cells
E
20 cells
F
30 cells
UMAP1
UMAP2

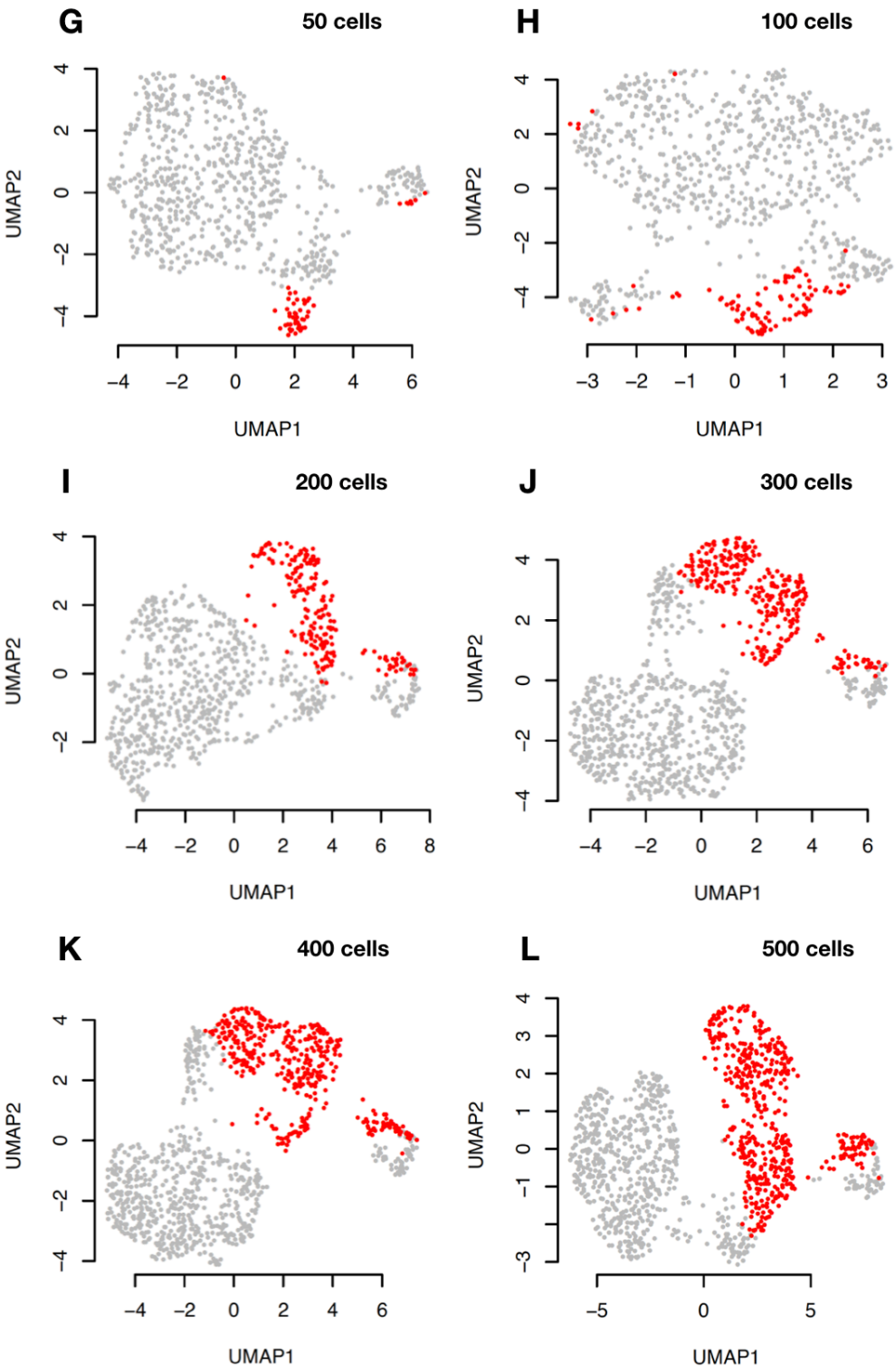
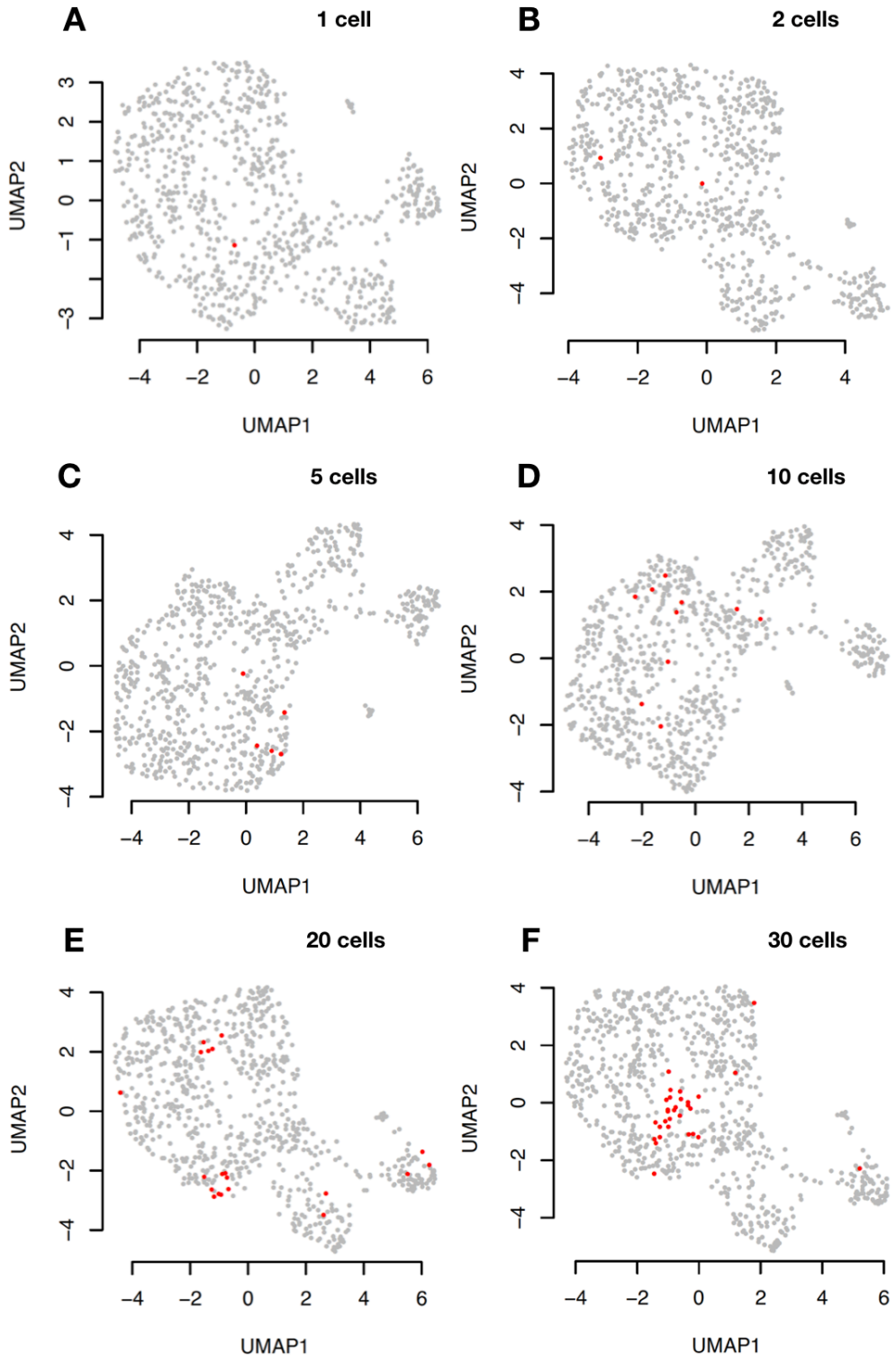


Figure 6.1: Subsampling analysis involving MCF-7 cells treated with niclosamide. Normalised gene expression data of MCF-7 cells treated with DMF (negative control) (grey) and MCF-7 cells treated with niclosamide (red), visualised using Uniform Manifold Approximation and Projection (UMAP) with dimensions = 20 for dimensionality reduction. Subsampling experiments involving 1 **(A)**, 2 **(B)**, 5 **(C)**, 10 **(D)**, 20 **(E)**, 30 **(F)**, 50 **(G)**, 100 **(H)**, 200 **(I)**, 300 **(J)**, 400 **(K)**, or 500 MCF-7 cells treated with niclosamide **(L)**. 683 MCF-7 cells treated with DMF were used for all analyses. Data reproduced with the permission of Xiaoli Ma, EPFL.


A
1 cell
B
2 cells
C
5 cells
D
10 cells
E
20 cells
F
30 cells
UMAP1
UMAP2

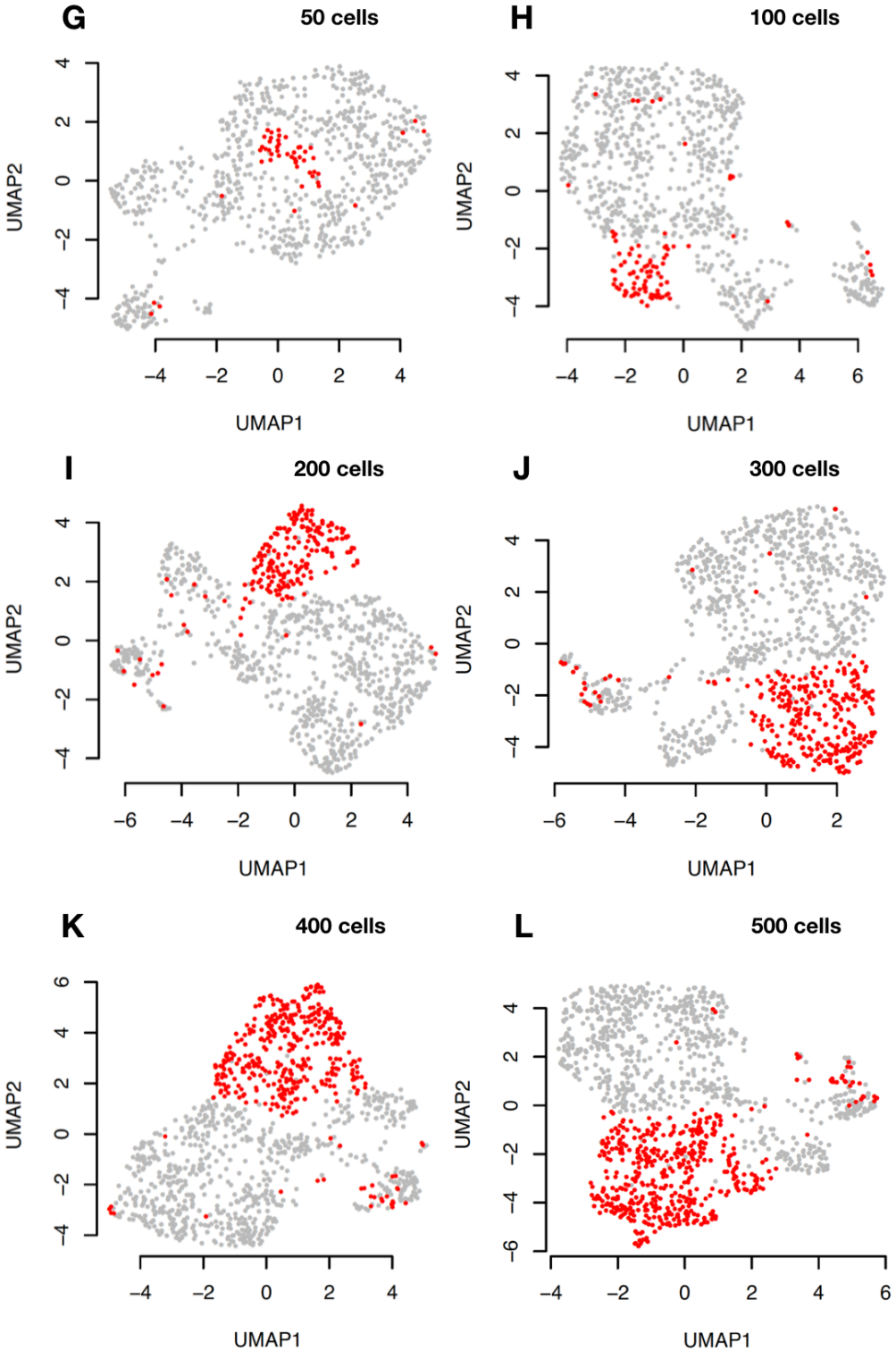


Figure 6.2: Subsampling analysis involving A375-P cells treated with niclosamide. Normalised gene expression data of A375-P cells treated with DMF (negative control) (grey) and A375-P cells treated with niclosamide (red), visualised using Uniform Manifold Approximation and Projection (UMAP) with dimensions = 20 for dimensionality reduction. Subsampling experiments involving 1 **(A)**, 2 **(B)**, 5 **(C)**, 10 **(D)**, 20 **(E)**, 30 **(F)**, 50 **(G)**, 100 **(H)**, 200 **(I)**, 300 **(J)**, 400 **(K)**, or 500 A375-P cells treated with niclosamide **(J)**. 763 A375-P cells treated with DMF were used for all analyses. Data reproduced with the permission of Xiaoli Ma, EPFL.

apparent when merely 30 - 50 cells are spiked in (Fig. 6.1F, G). In contrast, for A375-P cells treated with niclosamide, distinct clusters are only apparent when around 100 - 200 treated A375-P cells are added (Fig. 6.2H, I). This reinforce the idea that the exact number of cells required for the formation of one or more distinct clusters would be highly dependent on the effect size of the perturbation, which could vary between different perturbations and different cell lines.

Overall, it is clear that we need tens, possibly hundreds of cells, to detect a perturbation via clustering. This implies that with the current analysis method, perturbations cannot be detected on the single-cell level. The implications of this conclusion and alternative methods for the detection of perturbations are discussed in Chapters 7 and 12.

7 Discussion

In this part, we have assessed various potential model systems on their suitability to be utilised to establish a technology using single-cell transcriptomic analyses to study cell-cell interactions. We identified two model systems that are suitable, namely the system involving CHO-GLP1R-GFP and HEK293T-mCherry-Exendin-4 cells, and the system involving CD14-expressing U937 cells and 60bca hybridoma cells. The former involves cell-cell interaction via direct contact, while the latter involves the secretion of antibodies that interact with cell surface receptors. Therefore, the development and optimisation of our technology with both model systems will ensure that our technology remains widely applicable to different categories of cell-cell interactions. Consequently, we plan to utilise both these systems for subsequent optimisation experiments, as outlined in Section 7.1.

Nonetheless, we are also exploring a model system involving K-562 cells, lentiviruses and hybridoma cell lines secreting antibodies against the vesicular stomatitis virus glycoprotein (VSV-G). VSV-G is utilised by many widely-used lentiviral vectors, as it expands the range of host cells that can be transduced (Cronin et al., 2005). We have tested the supernatant from two hybridoma cell lines, namely 8G5F11 and IE9F9, for their ability to inhibit lentiviral transduction (Lefrancois and Lyles, 1982; Munis et al., 2018), and have demonstrated that incubation of lentiviral particles together with the 8G5F11 and IE9F9 hybridoma supernatants resulted in the inhibition of lentiviral transduction of K-562 cells (Appendix D, Fig. D.1).

We expect that lentiviral transduction will induce significant transcriptomic changes in transduced cells, that would be detectable by single-cell RNA-seq. In addition, we have cloned exogeneous genes (*OCT4*, SARS-CoV-2 nsp1) or shRNA against *GATA1* into the lentiviral transfer vectors, and we expect these genes or shRNA to further influence target cell transcriptomes in a manner detectable by emulsion-based single-cell transcriptomic analyses. Therefore, we expect a detectable transcriptomic difference between K-562 cells incubated with lentiviruses, in the presence or absence of the aforementioned anti-VSVG antibodies.

7.1 Future work

The next key experiment would be establishing a Drop-seq model screen with our chosen model systems. As explained in Chapter 5, three cell lines are required for each model system, namely the target cell line (CHO-GLP1R-GFP or CD14-expressing U937, in the presence of LPS), together with the stimulant cell line (HEK293T-mCherry-Exendin-4 or 60bca hybridoma cells) and a cell line genetically similar to the stimulant cell line, but that does not express the stimulant (WT HEK293T or H25B10 hybridoma cells, both of which act as negative controls) (Fig. 5.1). The stimulant cell line would be spiked into a population of negative control cells, prior to co-encapsulation with the target cell line, with the employment of different ratios of stimulant:negative cells revealing information on the sensitivity and specificity of our technology.

We expect to obtain single-cell transcriptomic data for the target cell lines, in conjunction with sequences that permit identification of the stimulant cell line, namely mCherry and Exendin-4, for the model system involving GLP1R and Exendin-4. For systems involving hybridoma cell lines, such as the system utilising U937 cells and the 60bca/H25B10 hybridoma cell lines, the sequences of interest would be those encoding the antibody heavy and light chains. We would be able to obtain these antibody sequences using the technology outlined in Part II.

This experiment will demonstrate if the technology can definitively identify droplets containing cells expressing activating factors that induce a specific transcriptomic effect in our target cells. It will also give us an idea of the efficiency of the technology, by telling us what percentage of positive droplets (those containing an activating cell together with a target cell) have a target cell transcriptomic profile distinct from droplets containing non-activated cells.

In addition, as Drop-seq will be utilised for our technology, we have to assess the sensitivity of Drop-seq in detecting our perturbations of interest, given that the 10X Genomics Single Cell 3' Gene Expression technology appears to outperform Drop-seq in sensitivity and precision (Ding et al., 2019; Zhang et al., 2019), and that some single-cell experiments were carried out with the 10X technology. As such, we may have to carry out additional Drop-seq experiments with target cells that have been activated in bulk, to study the ability of Drop-seq to detect transcriptomic changes at a single-cell level.

Subsequently, we would also like to test the performance of our technology at assessing cell-cell interactions in which one, or both, populations are heterogeneous. As most natural cell populations are heterogeneous, there is value in elucidating how usable the technology would be in analysing interactions involving heterogeneous cell populations. This could be done by utilising a diverse hybridoma library, such as the Precision Antibody hybridoma library that has been raised against

whole SK-BR-3 cells (as introduced in Chapter 3). The hybridoma library would act as a heterogeneous stimulant population, and would enable us to elucidate the sensitivity of our technology at detecting rare or infrequent cell-cell interactions. Naturally, these experiments are dependent on the bioinformatic analyses and on the identification of perturbations from single, or a small number of cells, which will be discussed further in Section 7.2.1.

7.2 Remaining challenges

The field of emulsion-based single-cell transcriptomics has advanced immensely in recent years. However, the utilisation of such technologies for the study of cell-cell interactions remain impeded, and this section will discuss the remaining challenges and theorise on how these may be overcome.

Three main challenges remain, namely:

1. Elucidating and reducing the minimum cell number to pick up perturbations
2. Overcoming Poisson distribution during cell and bead encapsulation
3. Establishing targeted sequencing

7.2.1 How many cells are needed to detect a perturbation?

The data in Chapter 6 suggests that between 30 - 100 cells would be required to detect a perturbation via clustering analysis. The analysis carried out was rather qualitative as clustering was determined by eye, and not in a quantitative manner. In addition, these analyses involved the random sampling of treated cells, suggesting that chance may have played a role in determining which cells were selected, and thus in where the threshold was. It would naturally be ideal to both repeat the random sampling and the analysis multiple times, and to develop a method to quantitatively grade the extent of clustering.

This analysis was carried out with MCF-7 and A375-P cells treated with niclosamide, a drug, and did not involve any perturbations that would be present within cell-cell interactions. As the effect size of the perturbation is expected to affect the number of cells required for perturbation detection, it would be ideal for us to carry out a similar analysis with a perturbation present in cell-cell interaction model systems, such as an antibody-based perturbation or with Exendin-4, which is involved in the interaction between CHO-GLP1R-GFP and HEK293T-mCherry-Exendin-4 cells.

In addition, the model screen mentioned above would be crucial. Specifically, the spike-in of known ratios of positive hits in a sea of negative events would give us a better measure of effect size and of the sensitivity of our technology, for our model system of interest. It could also be beneficial to utilise a system with

a maximal effect size, to truly minimise the number of cells required to detect a perturbation. Alternatively, the use of a model system that has almost-digital differential expression in one or more genes, for example, the expression of GFP in CHO-GLP1R-GFP cells, could improve the identification of perturbed cells and be utilised for the optimisation of analysis pipelines for the identification of single, or small numbers of perturbed cells from overall gene expression (i.e. gene expression excluding GFP).

This requirement for 30 - 100 cells to detect a perturbation severely limits the throughput of our technology, as it suggests that for a perturbation to be picked up, it must be represented in 30 - 100 droplets. This limits the number of perturbations, namely different stimulants, that can be investigated in a single experiment, as sequencing costs would rapidly become prohibitively high. This limitation on the throughput of the technology greatly reduces the attractiveness of the technology for wider use.

In addition, this implies that many identical stimulant cells would be required in every experiment, in order to detect a perturbation. These may be difficult to obtain, especially if the number of cells to be investigated in a single screen remains generally low, such that the stimulant cells of interest would have to form a significant part of the stimulant cell pool. For example, rare cell types would be difficult to investigate, as they are present at low percentages in cell populations. It also remains onerous to enrich for such rare cell types, especially as they remain poorly understood. Similarly, this would hinder the study of primary plasma B-lymphocytes, which have significant diversity at the single-cell level, with many individual clones forming a minuscule part of the whole cell population. Even within populations where cell expansion is viable, such as within a diverse hybridoma cell population, such processes would be both costly and time-consuming. Therefore, a necessary area of improvement would be to reduce the number of cells required to detect a perturbation.

Studies involving single-cell transcriptomic screens involving genetic perturbations are a good proxy for understanding how our question has been investigated and answered by others. As further explained in Section 1.6.8, these screens involve genetic perturbations, such as CRISPR inhibition or activation, with the perturbation conditions encoded by genetic markers within the same cell in which the perturbation occurs (Datlinger et al., 2017; Dixit et al., 2016; Hill et al., 2018; Xie et al., 2017).

Most of these techniques examined an average of 80 - 150 transcriptomes, i.e. single cells, per sgRNA, i.e. per perturbation. It was found that the sensitivity and specificity of detecting the correct genes regulated by the perturbation (as seen in bulk RNA-seq) were 80% and 90% respectively in Perturb-seq, when 100 single cells per guide were used (Dixit et al., 2016), but the authors suggested that a mere

tens of cells per perturbation would be sufficient to detect signature and state-level regulation. In CROP-seq, they estimated that 12 - 13 cells per perturbation would be adequate to obtain gene expression signatures congruent with their current data (Datlinger et al., 2017). In contrast, the power estimation carried out by Alda-Catalinas *et al.* suggested that approximately 400 cells were required per sgRNA in a CRISPRa screen, to identify Zygotic Genome Activation (ZGA)-like signatures in mouse embryonic stem cells (mESCs) (adjusted p-value < 0.00032, power of 0.8) (Alda-Catalinas et al., 2020). This reinforces the idea that the number of cells required is likely to be highly dependent on the effect size of the perturbation, and suggests that the figure of 30 - 100 cells from our data is within the expected range.

Similarly, the recent TAP-seq sensitivity analysis by Schraivogel *et al.* suggests that a minimum of 30 - 40 cells must be profiled per perturbation (Schraivogel et al., 2020). It must be noted that this conclusion was derived for targeted Perturb-seq, where only 74 genes were specifically examined. This was concluded from the use of well-established perturbations that specifically downregulated their known targets in cis, and resulted in wider detectable transcriptomic effects. This suggests that targeted sequencing, as will be further explored in a later subsection, could reduce the number of cells required to detect a perturbation.

Within the same analysis, the authors carried out parallel analyses with their Perturb-seq datasets (no targeted sequencing). They either considered all detected genes, or simply examined the 74 genes from the TAP-seq target panel, and found that the latter had a higher precision, as the analysis across all detected genes resulted in lower precision, due to an increase in the number of false-positive hits (Schraivogel et al., 2020). The authors then argued that this strengthens the argument for hypothesis-driven analyses in such experimental designs.

This is not unimaginable within our context, where the examination of a subset of highly differentially expressed and biologically relevant genes could enable more effective segregation of treated and untreated populations, perhaps enabling the identification of perturbed cells even when they are present in smaller numbers. This could be done by using genes for which we know would be differentially expressed, perhaps based on previous bulk or single-cell sequencing data. Naturally, this could have the implication that analyses would have to be hypothesis-driven, meaning that our analyses would no longer be target-agnostic, a loss of a selling point of our technology. Nonetheless, such an approach would still be useful for screening, namely for the identification of new antibodies that have the same effects as previously discovered antibodies, but potentially with differing effect sizes and/or that bind to different epitopes.

In addition, one could also envision the target-agnostic identification of outliers by examining highly differentially expressed genes, or by examining a subset of

genes associated with specific pathways of interest. For example, it could be possible to examine a subset of target genes that are linked to GPCR signalling, with the screening remaining target-agnostic in the context of individual specific receptors, which could be specifically useful in the study of orphan GPCRs. These possibilities will all require further work to establish and optimise.

Other pipelines have been developed for the analysis of single-cell CRISPR screening data, and it is not difficult to imagine utilising similar principles for the analysis of results from our technology. For example, the CRISPRa screen which seeked to identify regulators of the ZGA program in mESCs utilised MOFA+ (Argelaguet et al., 2020) for the characterisation of molecular signatures and the identification of hits in their screen (Alda-Catalinas et al., 2020). MOFA+ is capable of grouping cells by sgRNA (perturbation) and permitting the modelling of these groups with separate hyperparameters, to discover "factors" which could be linked to a subset of sgRNAs and could correspond to specific biological pathways of interest (Alda-Catalinas et al., 2020; Argelaguet et al., 2020). A similar analysis utilising stimulant sequences in place of sgRNAs can be easily envisioned.

Similarly, MUSIC (Model-based understanding of single-cell CRISPR screening) has been developed for the better elucidation of perturbation function in single-cell CRISPR screens. MUSIC utilises a computational framework based on Topic Models, which was traditionally developed for the study of semantics in text (Duan et al., 2019). Like documents about a specific topic contain certain words more or less frequently, gene expression profiles of single cells are expected to have higher or lower expression of certain genes, depending on the perturbation. In this case, a topic is a specific biological function that is associated with a group of differentially expressed genes, which permits each perturbed transcriptome to possess a proportion of memberships in different functional topics, rather than being classified in a cluster. The authors argue that this makes their analysis more sensitive in the detection of subtle phenotypic changes, when compared to traditional clustering. One could envision topic models being utilised for the characterisation of perturbations within our technology, where an increased sensitivity could reduce the number of cells needed for the detection of a single perturbation.

Yet, it must be noted that our biological question is somewhat different to those present in single-cell CRISPR screens based on transcriptomics. The single-cell CRISPR screens typically examine transcriptomic signatures to study the impact of the perturbation on specific pathways of interest, with a limited number of perturbations studied. This also comes with the implication that their definition of the minimum number of cells required to detect a perturbation would be rather stringent, as they would require a sufficiently low p-value, where they can be confident that the perturbation has a detectable effect on the transcriptome. In contrast, we would like to study a larger number of perturbations in a target-agnostic manner, and could accept slightly lower sensitivity (i.e. higher numbers of false

negatives), as we would simply like to detect a subset of perturbations that would be interesting. Nonetheless, while it is unlikely that the analysis pipelines used for single-cell CRISPR screens can be directly copied for our analysis, the principles utilised in these analyses do remain partially relevant and should be further explored as we seek to improve the detection of perturbations within our technology.

7.2.2 Regulating droplet occupancy

The second crucial problem that we have also considered earlier is that of the regulation of droplet occupancy. This is a crucial problem as next-generation sequencing is extremely costly. The processing of droplets containing only a single cell type is undesired, simply as the genetic material from these droplets are irrelevant for our analyses. This increases sequencing and reagent costs without contributing any useful data on cell-cell interactions, as explained in Chapter 4.

While we have developed a microfluidic workflow to selectively process droplets containing our desired cell occupancy (as mentioned in Chapter 4), this has only been reliably established for a single colour, and thus for a single cell type. Naturally, it would be helpful to have a higher level of control over the occupancy of droplets that are processed, either by selectively processing droplets with a desired occupancy, or by regulating what is encapsulated.

The former can be done by modifying our current workflow to detect multiple fluorescence signals. We would stain the stimulant and target cells with different cell dyes (such as Cell Trace CFSE and Cell Trace Violet) prior to encapsulation, and would utilise two lasers and their associated PMTs to detect the two different fluorescence signals. Picoinjection would be triggered only when fluorescence signals from both cell types are detected in a single droplet. This is not inconceivable, given that dual-colour droplet sorting for the selective enrichment of droplets containing two differently stained cells has previously been established in the lab (Hu et al., 2015).

Deterministic cell encapsulation has been previously described, where Dean's force was utilised to carry out inertial ordering of particles or cells to be encapsulated. Kenma *et al.* demonstrate an increased efficiency of single cell encapsulation, with 77% of droplets containing a single cell, which is significantly higher than the hypothetical Poisson limit of 37% (Kemna et al., 2012). The employment of multiple curved microchannels for deterministic cell encapsulation, coupled with electro-coalescence could enable the deterministic generation of droplets containing two cells of different cell types. Schoeman *et al.* have demonstrated such a workflow, with 40% of the droplets containing two cells after droplet fusion, although they did not distinguish between the cells encapsulated from the two different nozzles (Schoeman et al., 2014).

Other highly regulated single-cell encapsulation technologies have been described, for instance, those which deploy hydrodynamic traps to capture single cells prior to their encapsulation (Sauzade and Brouzes, 2017). While these could be fairly easily adapted to the capture and encapsulation of two cells, these techniques remain limited by their low throughputs.

The various possibilities outlined could improve our regulation of droplet occupancy or ensure that only droplets with our desired occupancies are further processed. However, it remains to be seen how effective these suggested improvements would be.

7.2.3 Targeted sequencing

With increasing numbers of cells to be analysed by single-cell RNA-seq, the costs increase rapidly, making large screens prohibitively costly. When coupled with the facts that only a subset of genes are variably expressed and that around 2% of genes consume over 50% of sequencing reads (Replogle et al., 2020), this supports the use of a targeted sequencing approach, to improve throughput and sensitivity without a correspondingly large increase in sequencing costs.

However, the use of a targeted sequencing approach implies that there must be a pre-selection of genes to be monitored. These genes can be associated with pathways or processes of interest, and as previously mentioned, a hypothesis-driven selection of genes may actually make it easier to identify cells of interest (Schraivogel et al., 2020). However, this prior knowledge may not be available for all applications, and it also makes any study no longer fully target-agnostic, with target agnosticism being a selling point of our technology.

Targeted sequencing can be carried out in emulsion-based single-cell transcriptomic technologies in various ways, as elaborated on in Section 1.6.6. The different techniques all seek to selectively enrich for specific sequences of interest, and this can be primarily done at three different steps of the emulsion-based single-cell RNA-seq protocol, namely during RNA capture, PCR amplification and enrichment via hybridisation capture.

Enrichment during RNA capture is facilitated by the presence of gene-specific oligonucleotide sequences that selectively capture mRNA of interest. This is easily realisable with the addition of gene-specific oligos during bead manufacture in InDrop (Zilionis et al., 2017). However, as we use Drop-seq in our application, DART-seq would be more appropriate (Saikia et al., 2019). Here, custom primers are enzymatically attached to a subset of poly-dT sequences on Drop-seq beads, to enable sequence-specific capture of a subset of mRNA.

Enrichment can also be carried out during PCR. This has been demonstrated by Constellation Drop-seq, which utilises a single primer per gene (Vallejo et al., 2019), and targeted Perturb-seq (TAP-seq), which utilises two different gene-specific

primers per gene (Schraivogel et al., 2020). These primers permit the specific amplification of cDNA corresponding to the mRNA of interest. Both techniques are fully compatible with Drop-seq, with the latter also compatible with the 10X Genomics Single-Cell 3′ Gene Expression platform.

Another example of PCR-based enrichment would be the antibody sequencing platform that we have established, which is outlined in Part II.

Enrichment is also possible via hybridisation capture to enrich for library fragments of interest. This is exemplified by direct-capture Perturb-seq (Replogle et al., 2020), which was only established for 10X Genomics gene expression libraries. Nonetheless, as hybridisation was carried out after library preparation, it is not difficult to envisage utilising the technology for a library prepared with a different emulsion-based single-cell RNA-seq technology.

We envision that the utilisation of targeted sequencing will enable the expansion of the possible throughput, without a proportionate increase in the cost of sequencing. Previously developed targeted single-cell transcriptomic technologies have demonstrated a reduction of 7- to 12-fold (Schraivogel et al., 2020) and 14-fold (Replogle et al., 2020) in the sequencing depth required for targeted single-cell transcriptomic sequencing to produce quantitatively similar data for the genes of interest, when compared to whole transcriptomic sequencing. This suggests that for the same sequencing cost, a switch from whole transcriptomic sequencing to targeted approaches could enable an increase in the number of single cells investigated by 6- to 13-times.

Overall, these three main challenges are closely associated with sequencing costs, which remain and will remain a significant determinant of the number of cells that can be processed in a single experiment, and thus the throughput of a single experiment. Any improvement in any of these three areas would improve the throughput of our technology for a given cost. Firstly, reducing the number of cells required to detect a perturbation of interest will enable the screening of more distinct perturbations with the same number of target cells sequenced. Improving the number of droplets containing the desired cell occupancy will reduce the amount of undesired sequencing data, increasing the amount of useful data, and potentially the number of cells that can be analysed with the same sequencing costs. Lastly, targeted sequencing will reduce the per-cell sequencing cost, given that only a subset of genes are monitored, enabling an increase in throughput for the same sequencing costs.

With the given experimental conditions, we could analyse approximately four thousand single cells in a single NextSeq run that would cost around € 2000. Given that our data in Chapter 6 demonstrate that a perturbation could only be detected with approximately 50 single cells, this works out to approximately 80 different conditions that could be detected and studied in such an experiment, making it

prohibitively expensive. This calculation was made from emulsion-based single-cell transcriptomic experiments without co-encapsulation, thus implying that the second challenge would not even be present under this circumstance, and that the actual throughput in an experiment involving co-encapsulation would be poorer. This suggests that improvements in the various areas are crucial in increasing the viability and attractiveness of the technology.

7.3 Future applications

The technology could be deployed for the screening of genetically encoded factors. One could envision screening factors for various biological processes, for example, for the identification of factors relevant to stem cell differentiation. Stimulant cells expressing different genetically encoded factors would provide different "niches" in droplets. By tracking the genetically encoded factor present, and the transcriptomic changes present in the target cells, factors of interest can be identified.

This illustrates the possibility of applying this technology, a cell-cell interaction technology that uses single-cell transcriptomics as a readout, to various biological systems and investigations.

Part II

A single-cell antibody sequencing workflow

8 A workflow for antibody sequencing

This part describes a novel modification to conventional Drop-seq (Macosko et al., 2015) that permits the sequencing of antibody heavy and light chains in a paired manner, while enabling the association of antibody sequences with whole transcriptome data from the same droplet. The workflow is introduced in this chapter, while two proof of concept analyses, namely the analysis of predefined hybridoma mixtures with known antibody heavy and light chain sequences and the analysis of a complex hybridoma mixture with unknown antibody heavy and light chain sequences, are presented in Chapter 9 and Chapter 10 respectively.

8.1 Introduction

Antibody sequencing is relevant both for advancing our understanding of the adaptive immune system, and in the identification and characterisation of antibodies in antibody screening. Antibody sequencing is complicated by the fact that the heavy and light chains of antibodies are encoded by different genes, with the endogenous heavy and light chain pairings being crucial for the reproduction and further study of antibodies of interest. This means that the pooling and bulk sequencing of antibody genes from antibody-secreting cells, which results in the loss of endogenous heavy and light chain pairings, is of limited usefulness.

As previously mentioned in Section 1.8.4, techniques that focus only on the sequencing of antibody variable heavy and light chains frequently utilise overlap extension PCR to physically link heavy and light chain sequences prior to sequencing or cloning for use in display libraries. These require the physical separation of antibody sequences derived from different cells, and frequently the physical separation of individual antibody-secreting cells, into individual reaction vessels. This has been carried out in nanowells (DeKosky et al., 2013) and in droplets (Adler et al., 2017a; Adler et al., 2017b; DeKosky et al., 2015; McDaniel et al., 2016; Rajan et al., 2018).

As explained in Section 1.8.5, the sequencing of antibody variable regions is not possible with unaltered conventional emulsion-based single-cell transcriptomic technologies (based on the 3′ capture of mRNA), due to the fact that the variable regions of the antibody genes are located nearer the 5′ end of the mRNA. The combination of fragmentation and short-read Illumina sequencing results in a failure to fully sequence the variable regions of the antibody-encoding transcripts in these technologies.

Other technologies have coupled the sequencing of antibody variable heavy and light chains with whole transcriptome analysis, either by capturing mRNA at their 5′ ends (10X Genomics 5′ Gene Expression Technology) (Goldstein et al., 2019; *Single Cell Immune Profiling*), by ligating gene-specific oligonucleotides onto Drop-seq beads to enable targeted capture of antibody-encoding mRNAs close to the variable region (DART-seq) (Saikia et al., 2019), or by sequencing 3′ labelled antibody-encoding cDNAs via long-read nanopore sequencing (Singh et al., 2019). These three techniques utilise different workarounds, but all permit the sequencing of antibody variable regions, together with their associated barcodes, to enable both the native pairing of heavy and light chain antibody sequences and the association of antibody sequences with the relevant single-cell whole-cell transcriptomic data.

Like the other strategies utilised for both sequencing antibody sequences and linking these to transcriptomic data, we aim to link cell barcodes with antibody heavy and light chain sequences, in a manner compatible with next-generation sequencing. Here, we introduce a method that is different from those previously outlined. Our method links cell barcodes and UMIs with antibody variable regions via an additional emulsion overlap extension PCR step, such that the resulting fragments are compatible with Illumina sequencing and both the antibody variable regions and their associated cell barcodes and UMIs can be sequenced. This would not only permit the association of the heavy and light chains, but would also enable association of the antibody sequences to the transcriptomic data, such that relevant antibody sequences can be associated with the relevant whole cell transcriptomes.

8.2 Experimental workflow

We have developed an emulsion overlap extension PCR workflow that is fully compatible with Drop-seq, to enable the sequencing of antibody heavy and light chain variable regions (Fig. 8.1). The experimental workflow has been primarily developed and optimised by Hongxing Hu, EMBL Heidelberg.

As outlined in Figure 8.1, we co-encapsulate antibody-secreting cells together with barcoded beads and lysis buffer, as outlined in Drop-seq (Macosko et al., 2015). The emulsion is then broken and subjected to reverse transcription and PCR. The PCR products are then encapsulated at a limiting dilution, to ensure that most

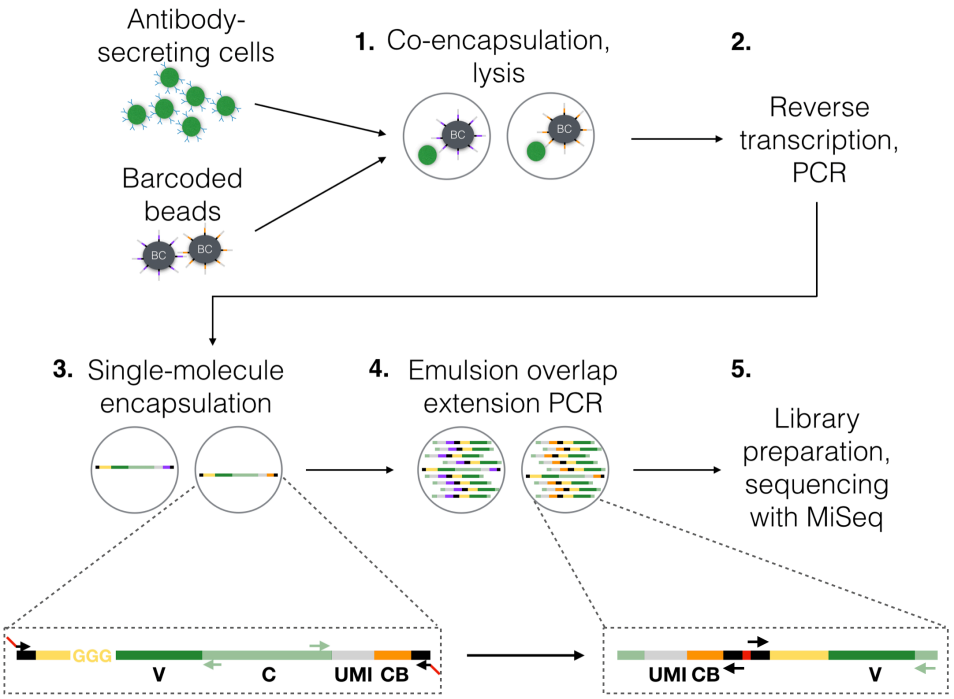


Figure 8.1: Workflow for Drop-seq-compatible antibody sequencing. The workflow for the antibody sequencing of antibody-secreting cells via the addition of small modifications to Drop-seq. **(1)** Antibody-secreting cells are co-encapsulated with barcoded beads and lysis buffer, to trigger cell lysis. **(2)** The emulsion is then broken and reverse transcription and PCR are carried out. **(3)** PCR products are encapsulated at a limiting dilution, such that a large majority of droplets contain either no PCR products or a single PCR product. DNA polymerase and four different groups of primers are also included in the droplets, which facilitate the emulsion overlap extension PCR to produce a fragment compatible with sequencing. **(Left inset)** Each DNA fragment with an antibody gene contains PCR handles (black), a segment corresponding to the template switch oligo (yellow, with "GGG"), the antibody variable (dark green, "V") and constant regions (light green, "C"), the UMI (grey, "UMI") and the cell barcode (orange, "CB"). The four primer types that are included in the droplets are pictured. The two that bind to the PCR handles are complementary (black arrows, with the complementary overhangs indicated in red), and the other two primers bind to the antibody constant region (light green arrows), permitting the specific amplification of antibody sequences. Two individual fragments are amplified from each template, but the use of primers with complementary overhangs results in the splicing and fusion of both fragments into a single fragment, as pictured in the right inset. **(Right inset)** The PCR product is pictured, with two flanking sequences corresponding to the primers that annealed to within the antibody constant regions (light green), the UMI (grey, "UMI"), the cell barcode (orange, "CB"), regions corresponding to the initial PCR handles (black) and the complementary overhang (red), the region corresponding to the initial TSO sequence (yellow) and the antibody variable region (dark green). The primers utilised for the sequencing of the UMIs and cell barcodes (black arrow, left), and for the sequencing of the antibody variable regions (black arrow, right and light green arrow) are also pictured. **(4)** Emulsion overlap extension PCR is carried out by thermocycling of the generated droplets. **(5)** A nested PCR is carried out to incorporate the relevant adaptors before purification and sequencing on a MiSeq. The fragments in the library would be comprised of an antibody heavy or light chain linked to the cell barcode and UMI (shown in the right inset). This means that the associated heavy and light chain sequences can be linked via the cell barcodes (purple, orange).

droplets contain no PCR products or only a single PCR product that would act as a template for the next PCR.

Reagents for the emulsion overlap extension PCR are also included in the droplets, namely DNA polymerase and four different types of primer (Fig. 8.1, left inset). Two groups of primers bind to the antibody constant region (Fig. 8.1, left inset, light green arrows), which permit PCR amplification to take place only in droplets containing an antibody-encoding mRNA. The other two primers bind to the flanking PCR handles and contain complementary overhangs (Fig. 8.1, left inset, black arrows with red overhangs).

Within each droplet containing an antibody fragment, two individual fragments would be amplified, one containing the antibody variable chain, and the other containing the UMI and cell barcode. However, the use of primers with complementary overhangs would result in the splicing and the fusion of both fragments into a single fragment (pictured in Fig. 8.1, right inset), that contains the antibody variable region linked to the UMI and cell barcode, with an altered orientation of the various components relative to each other.

The emulsion overlap extension PCR has to be carried out in droplets, as such a reaction in bulk would result in indiscriminate fusion of different barcodes with different antibody variable regions. Compartmentalisation introduced via the presence of droplets ensures that individual fragments are amplified and fused in independent reaction vessels, ensuring that the antibody variable regions are combined only with their associated cell barcode and UMI.

The resulting fragments are much shorter than the initial fragments, due to the exclusion of a large part of the antibody constant region. This makes the fragments short enough to be sequenced via paired-end Illumina sequencing.

8.3 Bioinformatic workflow

The bioinformatic pipeline has been developed by Charles Girardot and Jelle Scholtalbers, EMBL Heidelberg. Charles Girardot has also contributed significantly to the analysis of the data outlined in this Part of the thesis.

After sample demultiplexing, barcode detection and the demultiplexing of droplets are carried out. The paired-end reads, which correspond to reads from the two ends of each antibody variable region, are linked via the alignment of overlapping sequences present in the two reads, to give full-length antibody variable regions. Subsequently, these sequences are clustered first by UMI, to give a consensus sequence for each UMI, and then by droplet barcode, to give an idea of the identity and distribution of the different antibody chains in each droplet. IgBLAST is then carried out to identify the V, D, and J regions for each identified chain. For the light chains, only kappa chains, which are expressed in over 95% of

mouse primary B-cells, are analysed in the data presented in this thesis, although the technology has now been adapted for the analysis of both kappa and lambda chains (data not shown). More details of the bioinformatic analysis can be found in Section 13.6.

We have demonstrated the viability of our technology via the sequencing of various hybridoma mixtures. Firstly, we investigated various samples of hybridoma cells that were mixed at predetermined ratios, and for which we had the antibody heavy and light chain sequences for the component hybridoma cells. We also analysed a complex hybridoma mixture, for which we had no prior knowledge of the antibody heavy and light chains. The data for these two experiments are outlined in the subsequent chapters (Chapters 9 and 10).

9 Antibody sequencing of predefined mixtures of hybridoma cells

We first validated our technology using hybridoma mixtures containing hybridoma cells with known heavy and light chain sequences, which were mixed at predetermined ratios.

9.1 Experimental setup

In order to test our technology, we utilised four hybridoma cell lines (4D5, H25B10, anti-CD4 and OKT-9). We have previously sequenced the heavy and light chain sequences of these hybridomas via Sanger sequencing, and therefore used these sequences as reference sequences for the subsequent Illumina sequencing data.

We processed three samples containing hybridoma cells mixed at different ratios (Fig. 9.1). The hybridoma cells were sorted via FACS prior to the experiment, to remove dead cells. Two samples containing all four cell lines were prepared, namely at a ratio of 1:1:1:1, with all cells at the same concentration, at a ratio of 50:1:1:1, where anti-CD4 hybridoma cells were included at a 50-fold concentration compared to the other cell types. Additionally, a third sample (1:1:1+S) was prepared where 4D5, H25B10 and OKT-9 were mixed at a ratio of 1:1:1, and incubated for 20 minutes with media in which anti-CD4 hybridoma cells were previously cultured (anti-CD4 supernatant), to investigate the effect of free-floating RNA on our sequencing results.

The samples were processed as outlined in Section 8.2, with the heavy and light chains being amplified and sequenced separately. This gives us a total of six different samples, namely a heavy and a light chain sample for each of the three experimental conditions.

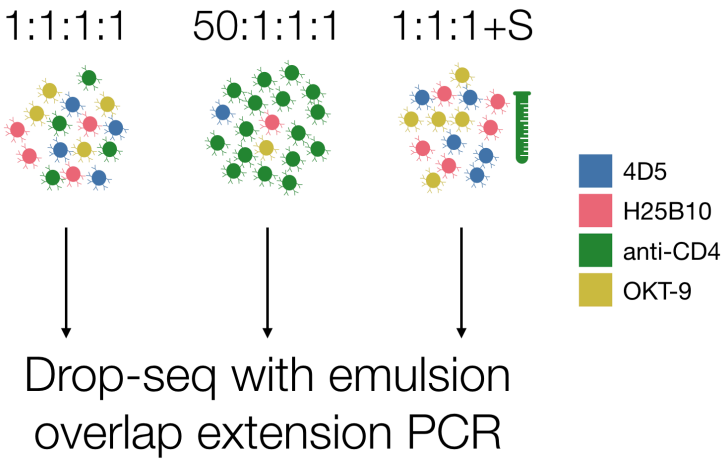


Figure 9.1: Experimental setup of hybridoma sequencing experiment. In the first sample (1:1:1:1), four hybridoma cell lines (4D5, H25B10, anti-CD4 and OKT-9) were mixed in equal proportion. In the second sample (50:1:1:1), three hybridoma cell lines (4D5, H25B10 and OKT-9) were mixed in equal proportion with anti-CD4 cells added at 50-fold the concentration of the other cell lines. In the third sample (1:1:1+S), three hybridoma cell lines (4D5, H25B10 and OKT-9) were mixed in equal proportion, and incubated with anti-CD4 supernatant for 20 minutes, prior to processing. All samples were then processed with the workflow outlined in Section 8.2. Hybridoma cells and supernatant of different origins are indicated with different colours (4D5: blue, H25B10: pink, anti-CD4: green, OKT-9: mustard)

9.2 Data processing

The data was processed as outlined in Sections 8.3 and 13.6. It must be noted here that for the data presented here, only kappa chains were detected. However, we expect this to have a negligible impact on the analysis, given that all the known light chain sequences of the four hybridoma cell lines are kappa light chains.

BLAST was used to align antibody heavy and light sequences against a custom database containing only the known antibody heavy and light chain sequences of 4D5, H25B10, anti-CD4 and OKT-9, which had been previously obtained with Sanger sequencing. This ensures that each antibody heavy or light chain obtained from next-generation sequencing, and each associated UMI, could be classified to a specific chain identity. The UMIs of the sequenced antibody chains were also associated with their respective droplet barcode, such that the antibody heavy or light chain identities of the various UMIs associated with each droplet were used to identify the hybridoma cell that was likely to have been encapsulated in each droplet.

This is visualised in Figure 9.2, where the number of UMIs associated with the different light chains for the different droplets in the 1:1:1:1 sample are visualised. The bar graphs for all six samples can be found in Appendix E.

Droplets containing five or fewer UMIs associated to identifiable chains specific to the library type (light chains for light chain libraries, heavy chains for heavy chain

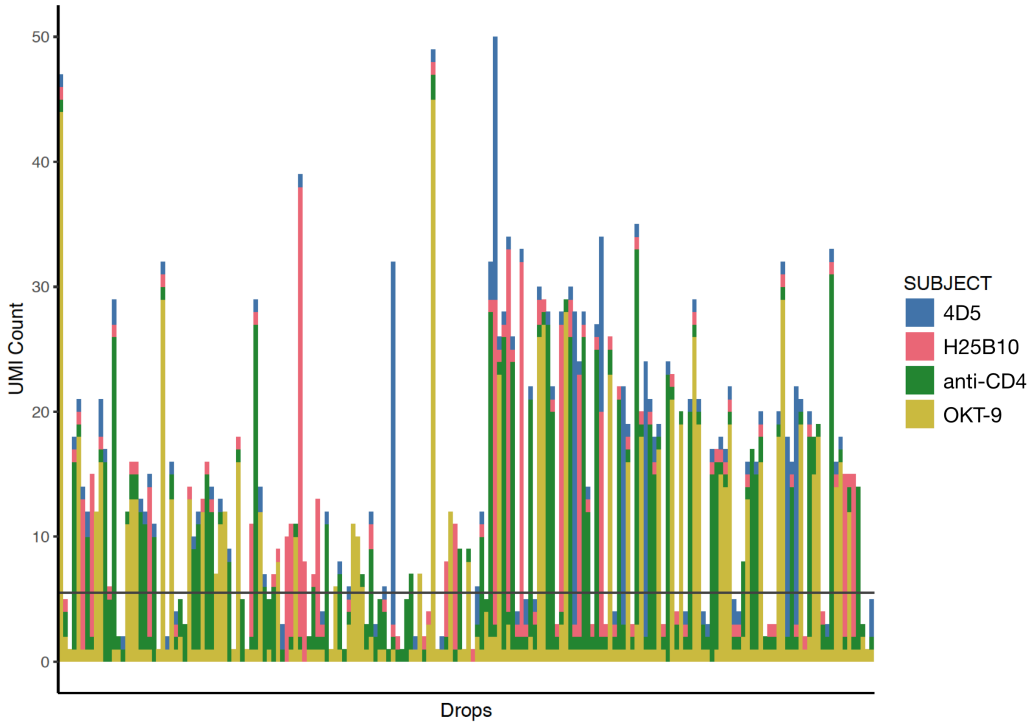


Figure 9.2: Distribution of UMIs associated with droplets for light chain sequencing of the 1:1:1:1 hybridoma mixture. Four hybridoma cell lines (4D5, H25B10, anti-CD4 and OKT-9) were mixed in equal proportion. Independent droplets are shown on the x-axis ("Drops") while the UMI count for each droplet is depicted on the y-axis ("UMI Count"). Each bar corresponds to the UMIs associated within a single droplet. The identity of the light chain associated with each UMI is indicated by the colour(s) of the bar ("SUBJECT", with 4D5: blue, H25B10: pink, anti-CD4: green, OKT-9: mustard). The grey line at UMI Count = 5.5 indicates the cut-off utilised. Droplets with bars above this cut-off were further processed, while droplets with bars below this, which have five or fewer UMIs associated to identifiable light chains, were rejected.

libraries) were rejected (Fig. 9.2), as we expected that droplets containing hybridoma cells would optimally have more detectable antibody sequences. From the remaining droplets, we classified droplets to specific hybridoma cells if more than half of their UMIs were associated with the same chain, and rejected the remaining droplets.

Between 86 and 488 droplets were detected for the six different samples (Fig. 9.3), with about 10% - 40% of droplets being discarded for most samples. For light chain sequencing, droplets that were discarded mostly had fewer than six UMIs (Fig. 9.3, grey), but for some heavy chain samples, there were more droplets with six or more UMIs in which no chain accounted for more than 50% of the UMIs (Fig. 9.3, black). The quality of the heavy chain library for 50:1:1:1 and the light chain library for 1:1:1:1 appear to be particularly poor, with comparably low UMI counts for each droplet and low numbers of droplets detected (Appendix E). Given that the heavy and light chain libraries were generated from the same samples, and that the three different samples for each chain (heavy or light) were processed together, no systematic reason for the variation in quality is readily apparent.

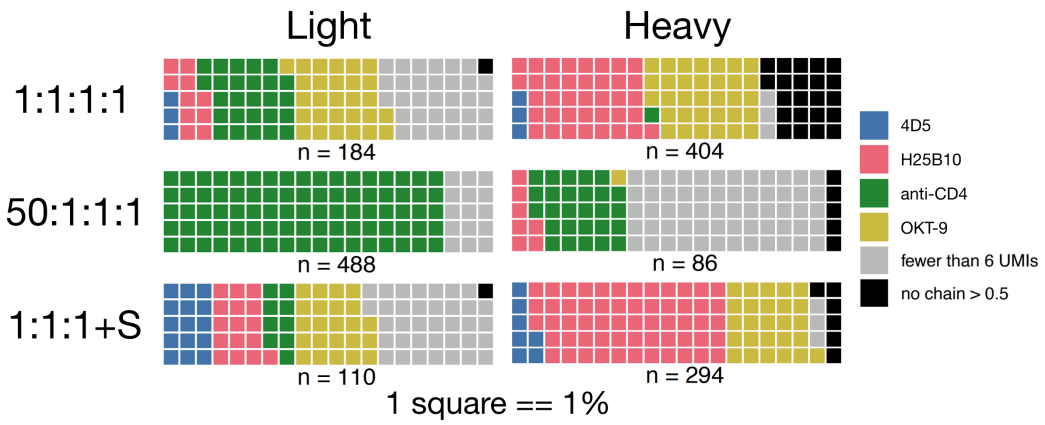


Figure 9.3: Identity of droplets for heavy and light chain antibody sequencing (including discarded droplets). The proportion of droplets rejected, or classified to each hybridoma cell line for both the heavy and light chain libraries of the three samples are visualised. The droplets assigned to specific hybridoma cell lines are indicated in colour (4D5: blue, H25B10: pink, anti-CD4: green, OKT-9: mustard), while the droplets that were rejected are indicated in grey (rejected as the droplet is associated with fewer than six UMIs) or black (rejected as no single chain accounts for more than half of the UMIs). Each square represents 1% of droplets in a sample, with the total number of droplets (n) for each sample noted below each waffle plot.

9.3 Results

The distribution of the droplets and their respective heavy and light chain identities, after the removal of the discarded droplets, are shown in Figure 9.4. While the ratios detected are not exactly representative of the ratios of the different cell types mixed, the general trend is coherent with the experimental setup.

When comparing the 50:1:1:1 and 1:1:1:1 samples, where the former had a concentration of anti-CD4 hybridoma cells that was 50-fold compared to the other cell types, a corresponding increase in the number of droplets assigned to anti-CD4 (green) is evident (Fig. 9.4). This is apparent for the light chain libraries, where anti-CD4 makes up about 39% of droplets (48/124 assigned droplets) in the 1:1:1:1 sample, which increases to over 98% (406/414 assigned droplets) in the 50:1:1:1 sample. Similarly, for the heavy chain samples, anti-CD4 constitutes about 1.6% of droplets (5/304 assigned droplets) in the 1:1:1:1 sample, which increases to over 76% (23/30 assigned droplets) in the 50:1:1:1 sample. A corresponding proportional decrease is also noticeable for the antibody chains of the other cell lines in the 50:1:1:1 sample, as compared to the 1:1:1:1 sample.

In contrast, the absence of anti-CD4 hybridoma cells, in the presence of anti-CD4 supernatant, resulted in a reduction in the number of droplets assigned to anti-CD4 (green) (Fig. 9.4). This is perceivable for the light chain libraries, where anti-CD4 accounts for about 39% of droplets (48/124 assigned droplets) in the 1:1:1:1 sample, which decreases to about 14% (10/69 assigned droplets) in the 1:1:1+S sample. Similarly, for the heavy chain samples, anti-CD4 represents about 1.6% of droplets

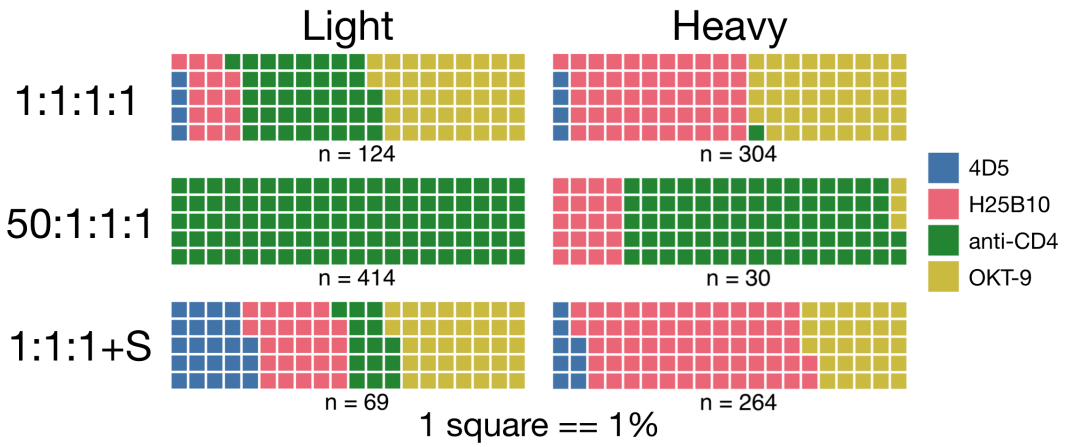


Figure 9.4: Identity of droplets for heavy and light chain antibody sequencing (excluding discarded droplets). The proportion of droplets classified to each hybridoma cell line, excluding the rejected droplets, for both the heavy and light chain libraries of the three samples are visualised. The identities of the droplets are indicated in colour (4D5: blue, H25B10: pink, anti-CD4: green, OKT-9: mustard). Each square represents 1% of droplets in a sample, with the total number of droplets (n) for each sample noted below each waffle plot.

(5/304 assigned droplets) in the 1:1:1:1 sample, which decreases to a mere 0.4% (1/264 assigned droplets) in the 1:1:1+S sample. A corresponding proportional increase is also clearly seen for the antibody chains of the other cell lines in the 1:1:1+S sample, as compared to the 1:1:1:1 sample.

However, while the percentage of droplets assigned to anti-CD4 (green) is smaller in the 1:1:1+S sample than in the 1:1:1:1 sample (Fig. 9.4), there are still droplets that are assigned to anti-CD4, suggesting that free-floating RNA present in the hybridoma supernatants do indeed contribute to contamination to a small extent.

This is also supported by the bar graphs for the different samples (Fig. 9.2 and Appendix E), where it is evident that many droplets contain a few antibody chains from cell lines other than the cell line that they are assigned to. Given that cells are encapsulated at a low concentration, such that most cell-containing droplets would contain only a single cell, this contamination is likely to arise from free-floating RNA in the supernatant. However, the problem is not a serious one as a majority of droplets can still be confidently assigned to one hybridoma cell line.

The ratios of the antibody chains from the different hybridoma cell lines vary between the heavy and light chain libraries, even for the same samples (Fig. 9.4). This could indicate differences in capture or amplification efficiency of the antibody heavy and light chains, or differences in heavy and light chain expression, even within each hybridoma cell line.

Additionally, there is clearly a variation in the efficiency of picking up heavy or light chains from different hybridoma cell lines. This is particularly evident

when considering the 1:1:1:1 sample. Here, cells from four different hybridoma cell lines were mixed in equal proportion, but the resulting droplets were not equally distributed across the four cell lines (Fig. 9.4). Instead, there is an over-representation of OKT-9 and an under-representation of 4D5, in both the light and heavy chain libraries, as well as an under-representation of anti-CD4 in the heavy chain library. This may be attributed to differences in antibody mRNA expression across different cell lines. In particular, the low numbers of droplets assigned to 4D5 could be associated with the fact that 4D5 does not produce productive heavy chains (Fig. 5.3), and that the antibody concentration of 4D5 hybridoma supernatant is generally over 85x lower than that of other hybridoma cell lines (Fig. 5.2), suggesting that 4D5 cells may express particularly low levels of antibody mRNA (Section 5.1).

Overall, we have generated the first proof of concept for our Drop-seq-compatible antibody sequencing workflow. We sequenced samples comprising of mixes of four different hybridoma cell lines at different ratios, and demonstrated that the heavy and light chains of all four hybridoma cell lines could be detected, sequenced and matched to previously known sequences. In addition, droplets could be assigned to the different hybridoma cell lines, with the proportion of droplets assigned to the different cell lines changing in a logical manner across the different samples, based on differing proportions of cells processed in the different samples. This indicates that we are capable of detecting antibody-secreting cells present at different proportions, with our data providing information on the relative proportion of the different cells present.

10 Antibody sequencing of a complex hybridoma mixture

We next employed our technology to investigate a complex hybridoma mixture, where we had no prior knowledge of the sequences of the antibody heavy and light chains.

10.1 Experimental setup

As previously mentioned in Chapter 3, hybridoma cells that secrete antibodies that bind whole SK-BR-3 cells were generated by Precision Antibody. These were generated utilising primary B-cells from mice that were shown to have an immune response against SK-BR-3 cells (Fig. 3.1B). Thus, we expect that a subset of these hybridoma cells would secrete antibodies that would bind to membrane receptors present on SK-BR-3 cells and potentially induce transcriptomic changes. As this is a model system under investigation, we were also interested to appraise the ability of our technology to analyse the antibody diversity within this hybridoma sample.

The samples were processed as outlined in Section 8.2, with the heavy and light chains being amplified separately but sequenced together. Hybridoma cells from a single flask were split into two identical groups prior to Drop-seq processing and sequencing. This gives us a total of four different samples, namely a heavy and a light chain sample for each group.

It must be noted here that mouse immunoglobulins have either the kappa or the lambda light chain, with individual B-lymphocytes exclusively expressing only one class of light chain. Our analysis pipeline only detects kappa chains, but we expect this to have a minimal impact on the analysis, given that over 95% of primary mouse B-cells express kappa light chains, while less than 5% express lambda light chains.

10.2 Results

As in the previous experiment, the UMIs of the sequenced antibody chains were associated with their respective droplet barcodes, but no alignment to known heavy or light chains was carried out.

We detected over 700 droplet codes for each group, of which over 20% could be associated with both heavy and kappa light antibody chains (Fig. 10.1, blue). This figure is suboptimal, and suggests that our sampling is insufficient to pick up paired heavy and kappa light chains for 80% of droplets, which would be required for a complete understanding of the detected antibodies, and for any subsequent cloning and reconstruction of antibody sequences of interest.

A majority of the remaining droplets were associated with only antibody heavy (Fig. 10.1, pink) or kappa chains (Fig. 10.1, green), with almost no droplets associated with neither, which is expected, as the workflow involves sequence-specific amplification of antibody heavy and kappa chains.

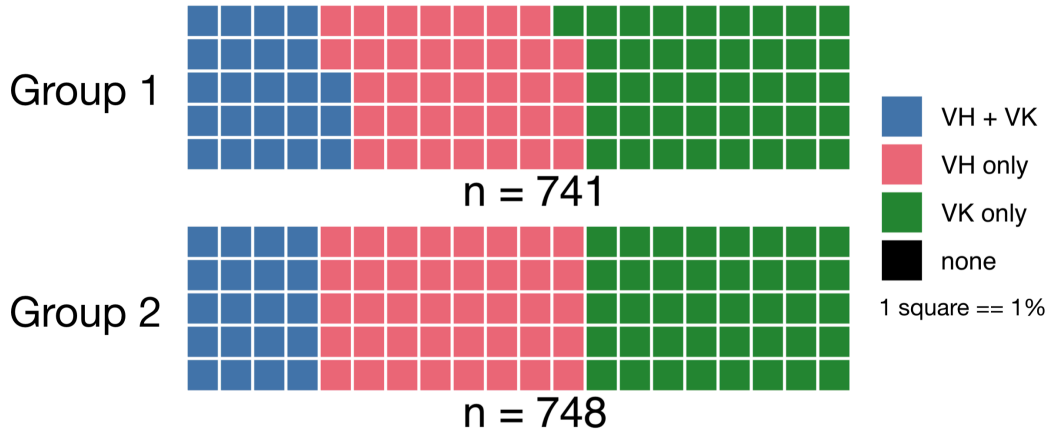


Figure 10.1: Association of V_H and V_K to droplets. The proportion of droplets that have both an associated heavy (V_H) and kappa chain (V_K) (VH + VK, blue), only an associated heavy chain (V_H) (VH only, pink), only an associated kappa chain (V_K) (VK only, green) or neither a heavy or a kappa chain (none, black). Each square represents 1% of droplets in a sample, with the total number of droplets (n) for each group noted below each waffle plot.

When considering the droplets that were associated with both heavy and kappa antibody chains (Fig. 10.1, blue), there are large numbers of unique V_H and V_K pairs for both groups (Group 1: 152, Group 2: 135), which both correspond to 89% of the number of droplets for which V_H and V_K pairs were detected (Table 10.1).

Group	Number of droplets with V_H/V_K pairs	Number of unique V_H/V_K pairs	Percentage
Group 1	171	152	89%
Group 2	152	135	89%

Table 10.1: V_H/V_K pairs.

An examination of the heavy and kappa chains detected for each library revealed that while over 400 heavy or kappa chains were detected within each library, the number of unique chains was much lower (Table 10.2). Specifically, the numbers of unique heavy chains were 188 and 181 for Groups 1 and 2 respectively, which correspond to 43% and 40% of the number of detected heavy chains. Similarly, the numbers of unique kappa chains were 144 and 135 for Groups 1 and 2 respectively, which is equal to 30% of the number of detected kappa chains for both groups.

Library	Number of chains	Number of unique chains	Percentage
Group 1, heavy chains	434	188	43%
Group 2, heavy chains	447	181	40%
Group 1, kappa chains	473	144	30%
Group 2, kappa chains	453	135	30%

Table 10.2: Unique V_H and V_K chains.

The exact distribution of the unique heavy and kappa chains across droplets has been visualised in Figures 10.2 and 10.3. While 57% (107/188) and 52% (94/181) of the unique heavy chains identified in Group 1 and Group 2 respectively appeared in only one droplet each, over 20 heavy chain sequences for each group were found in five or more droplets (Fig. 10.2). Similarly, while 44% (64/144) and 37% (70/135) of unique kappa chains identified in Group 1 and Group 2 respectively were present in only one droplet each, over 25 kappa chain sequences for each group were found in five or more droplets (Fig. 10.3).

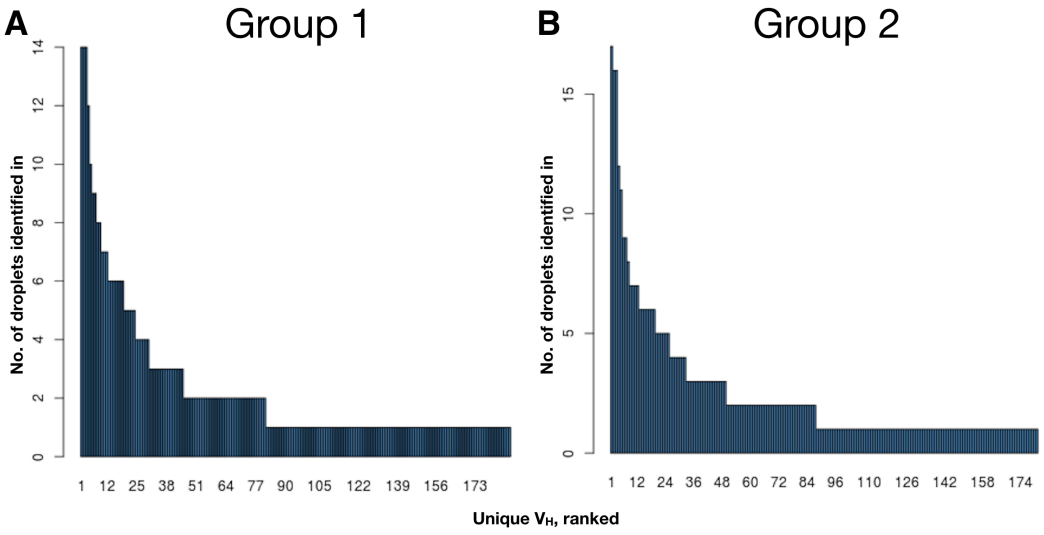


Figure 10.2: Occurrences of unique heavy chains in droplets. The unique heavy chains have been ranked by the number of droplets they are found in, and plotted along the x-axes, while the y-axes indicate the number of droplets in which each chain has been identified in. Data is shown for the heavy chain libraries from Group 1 (**A**) and Group 2 (**B**). Data reproduced with the permission of Charles Girardot, EMBL Heidelberg.

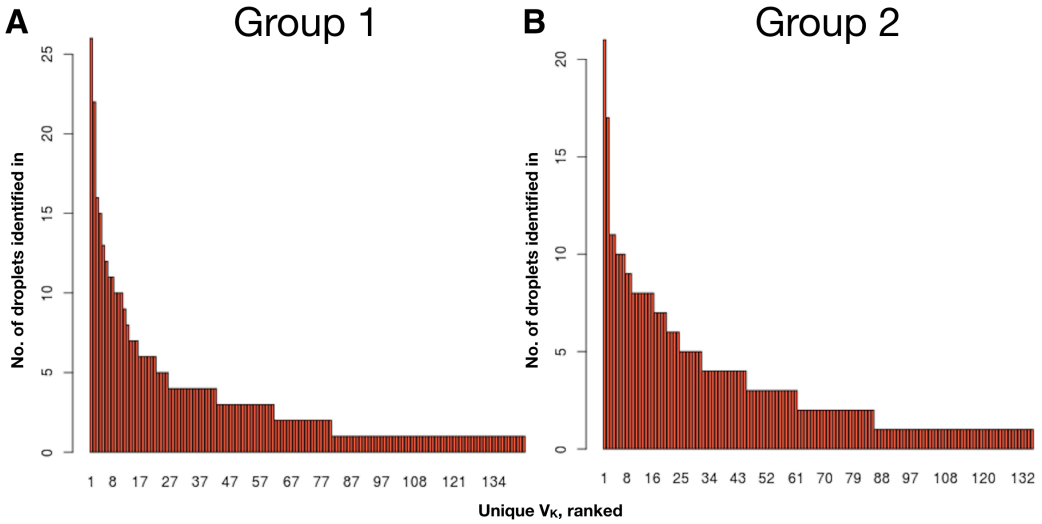


Figure 10.3: Occurrences of unique kappa chains in droplets. The unique kappa chains have been ranked by the number of droplets they are found in, and plotted along the x-axes, while the y-axes indicate the number of droplets in which each chain has been identified in. Data is shown for the kappa chain libraries from Group 1 (**A**) and Group 2 (**B**). Data reproduced with the permission of Charles Girardot, EMBL Heidelberg.

There is a reasonable level of congruence between the frequently detected heavy and kappa chains between Group 1 and Group 2. For the heavy chains, of the top 10 most frequently occurring chains in Group 1 and Group 2, three chains are present in both libraries. For the kappa chains, of the top 10 most frequently occurring chains in Group 1 and Group 2, seven chains are present in both libraries (Table 10.4, Original Rankings).

However, a closer examination of the data suggests that some of the over-represented kappa chains may have arisen from PCR errors or errors on the bead barcodes, rather than being due to a biological lack of diversity in the hybridoma population (Table 10.3). This can be seen by examining plots in which the barcodes associated with a particular heavy or kappa chain antibody sequence are plotted (Fig. 10.4 and 10.5).

Library	Unique chains with ≥2 barcodes	Unique chains with likely PCR errors	Unique chains with likely bead errors	Unique chains without errors
Group 1, heavy chains	81	5	1	75/81 (93%)
Group 2, heavy chains	87	10	0	77/87 (89%)
Group 1, kappa chains	80	5	3	72/80 (90%)
Group 2, kappa chains	85	9	3	73/85 (86%)

Table 10.3: Errors in barcode counts of V_H and V_K chains. The number of unique chains associated with two or more barcodes were quantified. Errors were defined as barcodes pairs or groups that had only one or two nucleotide differences. Single nucleotide differences located throughout the barcode were classified as likely PCR errors, while nucleotide differences that occurred were located at the end of the barcode and in groups of 4 or 16 were classified as likely bead errors.

The examination of the most frequently detected heavy chains from Group 1 and

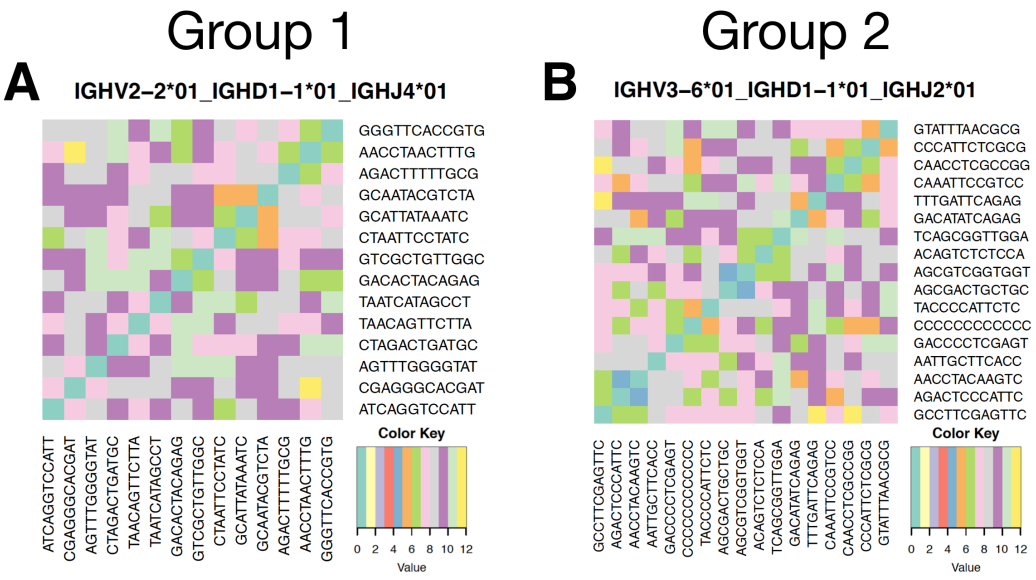


Figure 10.4: Analysis of droplet code distances for the most frequently detected heavy chains. The droplet barcodes of the most frequently detected heavy chains in Group 1 (**A**, n = 14) and Group 2 (**B**, n = 17) were examined. Droplet barcodes are found under and to the right of the heatmap, while the different colours indicate the number of nucleotide differences between pairs of barcodes. The germline V, D, and J sequences of the antibody chain are indicated in the title. Data reproduced with the permission of Charles Girardot, EMBL Heidelberg.

Group 2 reveal that the various barcodes have differences of at least five nucleotides, suggesting that they did indeed arise from different droplets (Fig. 10.4). This is also the case for most detected heavy chains, with the occasional presence of pairs of barcodes which differ by a single nucleotide (Fig. F.1, Appendix F). While mutations present within droplet barcodes may cause a small overestimation in barcode number for some heavy chains, this has been noted for only a small number of unique heavy chains (5/81 for Group 1, 10/87 for Group 2) (Table 10.3), with most identified events only involving two barcodes. Overall, as 93% and 89% of unique heavy chains with two or more associated barcodes are not problematic (Table 10.3), we are confident that the errors do not pose a significant problem for the identification and quantification of heavy chains.

Similar analyses can be conducted for the kappa chain sequences. Scrutiny of the most frequently detected kappa chains from Group 1 (Fig. 10.5A) and Group 2 (Fig. 10.5C) revealed that a majority of detected barcodes are highly similar. For the most frequently detected kappa chain from Group 1, 16 barcodes (bottom left corner, pale yellow and light grey cyan) differ by only one or two nucleotides (Fig. 10.5A). Similarly, for the most frequently detected kappa chain from Group 2, 16 barcodes (top right corner, pale yellow and pale purple) differ by only one or two nucleotides (Fig. 10.5C). The fact that the differences occur exclusively in the last two nucleotides of the barcode, with all four possible nucleotides present, suggest that they may have arisen from errors in bead barcoding, where the barcodes are shorter than expected,

with the last two nucleotides being part of the UMIs. This suggests that the true number of barcodes for the kappa chains IGKV1-117*01_IGKJ1*01 (Fig. 10.5A) and IGKV8-27*01_IGKJ5*01 (Fig. 10.5C) should be 11 and 6 barcodes respectively.

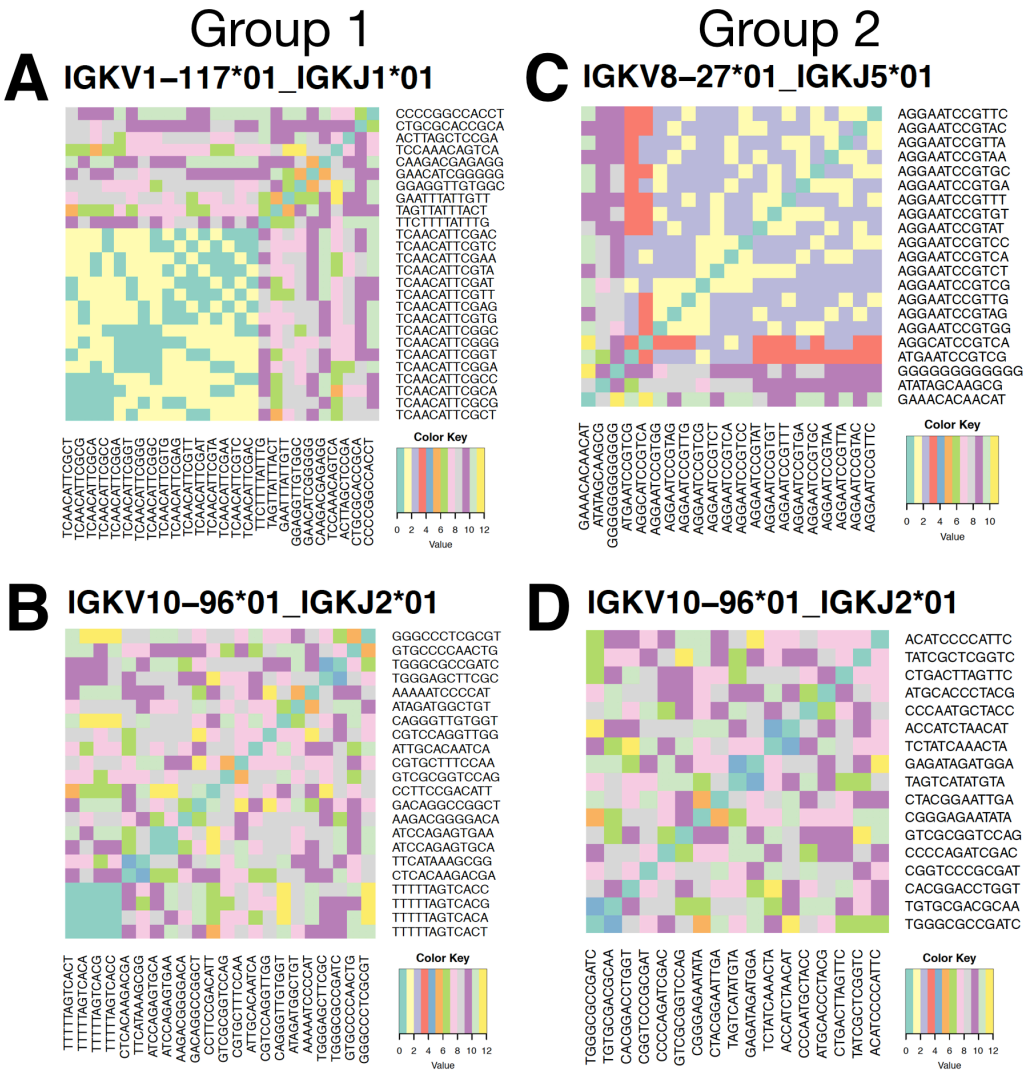


Figure 10.5: Analysis of droplet code distances for the top and second most frequently detected kappa chains. The droplet barcodes of the most and second most frequently detected kappa chains in Group 1 (**A**: most (n = 26) and **B**: second most (n = 22)) and Group 2 (**C**: most (n = 21) and **D**: second most (n = 17)) were examined. Droplet barcodes are found under and to the right of the heatmap, while the different colours indicate the number of nucleotide differences between pairs of barcodes. The germline V and J sequences of the antibody chain are indicated in the title. Data reproduced with the permission of Charles Girardot, EMBL Heidelberg.

A similar issue has also been noted for the second most frequently detected kappa chain in Group 2, where four barcodes (bottom left, light grey cyan) differ from each other solely in the last nucleotide (Fig. 10.5B). A thorough analysis of all the unique kappa chains reveals that three unique kappa chains each in Group 1 (3/80) and Group 2 (3/85) have similar issues (Table 10.3). In addition, single nucleotide differences between pairs of barcodes, which are more likely to have

arisen during PCR, were noted for 4/80 and 9/85 kappa chains for Group 1 and Group 2 respectively (Table 10.3). This suggests that 90% (72/80) and 86% (73/85) of unique kappa chains from Group 1 and Group 2 respectively are not problematic and have accurate barcode counts, including the second most frequently detected kappa chain from Group 2 (Fig. 10.5D), and that these errors do not drastically alter the conclusions.

	Original Rankings		**Corrected Rankings**	
Germline V and J (kappa chains)	**Group 1**	**Group 2**	**Group 1**	**Group 2**
IGKV1096*01_IGKJ2*01	2	2	1	1
IGKV14126*01_IGKJ2*01	4	4	2	3
IGKV37*01_IGKJ2*01	5	3	3	2
IGKV1096*01_IGKJ1*01	6	15	4	10
IGKV1117*01_IGKJ1*01	1	8	5	6
IGKV17121*01_IGKJ2*01	7	10	6	13
IGKV620*01_IGKJ2*01	8	19	7	18
IGKV819*01_IGKJ2*01	9	5	8	4
IGKV830*01_IGKJ1*01	10	7	9	5
IGKV830*01_('IGKJ1*01', 'IGKJ1*02')	11	9	10	7
IGKV1993*01_IGKJ1*01	3	16	13	11
IGKV625*01_IGKJ1*01	16	6	16	12
IGKV1110*01_IGKJ1*01	21	14	21	9
IGKV1110*01_IGKJ2*01	59	12	59	8
IGKV827*01_IGKJ5*01	NA	1	NA	19

Table 10.4: Kappa chain rankings based on barcode frequencies. Kappa chains were ranked based on the number of associated barcodes, with the chains with the highest barcode frequencies ranked more highly. All chains that were in the top ten, in either the original or corrected rankings, are shown. The corrected rankings were derived by assuming that all barcodes with only one or two differences were derived from the same barcode. "NA" indicates where the kappa chain sequence was not detected for two or more barcodes.

As we can identify these errors, they should ideally be excluded from subsequent analyses. The correction of these errors results in data that appears more congruent across the different groups. This is exemplified by the fact that the second most frequently detected kappa chains for Group 1 and Group 2 are the same (IGKV10-96*01_IGKJ2*01, Fig. 10.5B, D). Given that the highly similar barcodes present in the most frequently detected kappa chains for Group 1 and Group 2 are likely to be derived from the same bead, and that the true number of barcodes for the kappa chains IGKV1-117*01_IGKJ1*01 (Fig. 10.5A) and IGKV8-27*01_IGKJ5*01 (Fig. 10.5C) should be 11 and 6 barcodes respectively, what we termed the second most frequently detected kappa chain (IGKV10-96*01_IGKJ2*01, Fig. 10.5B, D) would actually be the most common kappa chain sequence for both groups, with 18 and 21 unique barcodes for Group 1 and Group 2 respectively (Table 10.4). With the corrected rankings, the top three kappa chains in Group 1 are the same three sequences identified as the top three kappa chains in Group 2. In addition, of the top 10 most frequently occurring chains in Group 1 and Group 2, eight chains are

present in both libraries, which is up from seven in the original rankings (Table 10.4).

Overall, these analyses suggest that some of the high barcode counts for some antibody chains have arisen from PCR or barcoding errors. However, other high barcode counts appear to be biological, suggesting that the complex hybridoma library is not as diverse as expected, with a subset of heavy and kappa chains present in multiple droplets. We cultured the Precision Antibody hybridoma cells for one week prior to the experiment, and we hypothesise that this time period may have been too long, such that certain hybridoma populations may have expanded. Further optimisation of the culture process prior to antibody sequencing can be carried out, to ensure that cells are of sufficiently good quality, while minimising the loss of diversity that may occur upon extended culture.

11 | Discussion

We have generated the first proofs of concept for our Drop-seq-compatible antibody sequencing workflow. We sequenced samples comprising of mixes of four different hybridoma cell lines at different ratios, and demonstrated that the heavy and light chains of all four hybridoma cell lines could be detected. We detected various ratios of different heavy and light chains which were coherent with the ratios at which the cells were included in the various samples. We have also sequenced a complex hybridoma library where we had no prior knowledge of the antibody heavy and light chains, demonstrating that our technology can detect antibody heavy and light chain sequences from a diverse library.

Our technology is dependent on and compatible with Drop-seq, and thus permits the simultaneous study of antibody heavy and light chain sequences and of whole transcriptomes from the same samples. The association of the antibody heavy and light chains with their corresponding cell barcodes would enable the association of antibody chain information with the specific single-cell transcriptomes with the same cell barcodes.

There are conspicuous opportunities for improvement. We have noted the presence of contaminating chains in most droplets, and hypothesise that these arise from free-floating RNA, which may in turn arise from dead cells present within the samples. While the cells utilised in the experiment were sorted by fluorescence-activated cell sorting (FACS) to remove dead cells, and were washed and kept at low temperatures throughout the processing, it is possible that a small proportion of cells are dying during the process and releasing mRNA. Further optimisation of cell handling could mitigate this issue.

We have also observed that our sampling was insufficient to pick up both heavy and light chains for many droplets. This could potentially be overcome by increasing the number of fragments sequenced, or by processing a smaller number of STAMPs at each time, to ensure that more pairs of associated antibody heavy and light chains would be detected, but further experiments or *in silico* analyses are required to assess the suitability of these measures.

We have also noted that the full length heavy chain variable sequences were occasionally slightly too long for an alignment to be generated by alignment of the

paired reads. In most of these cases, it is still possible to obtain the CDR3 region, the region crucial for the determination of antigen specificity of the antibody, from the second read. While this is less desirable than obtaining the full length heavy chain read, the CDR3 region could still provide useful information and enable cloning of a key region of the antibody heavy chain variable region. This should be further explored and evaluated.

For the two datasets presented, only kappa light chains were detected, and lambda chains were not included within the analysis. This is not a problem in the experiment involving mixes of four different hybridoma cell lines, as all four of these hybridoma cell lines express only kappa light chains. In contrast, while this could have resulted in the exclusion of a small number of lambda chains in the complex hybridoma mixture, the extent of the problem is expected to be small, give that over 95% of primary mouse B-cells express kappa light chains, while less than 5% express lambda light chains. Nonetheless, our analysis pipeline should be, and has already been expanded to include the analysis of lambda chains (data not shown).

Our results have also identified a potential vulnerability. Specifically, it appears that some barcoded beads may have shorter cell barcodes, which can cause errors in our analysis and had resulted in an over-estimation of the number of barcodes associated with specific antibody chains. This is not surprising, given that previous analyses have found that about 10% of Drop-seq beads contain a one-base deletion in cell barcodes (Zhang et al., 2019). It has also been noted by the creators of Drop-seq that a batch of barcoded beads generated before June 2015 has a subset of barcodes (about 10 - 20%) that shared the first 11 bases but that differed at the last base (Nemesh, 2018). Extra caution is required during data analysis to filter out these barcodes. They can be excluded by the tolerance of a mismatch of two nucleotides for cell barcodes, or by implementation of the DetectBeadSynthesisErrors package, which should detect and repair barcode indel synthesis errors (Nemesh, 2018).

More work should also be done to quantitatively estimate the sensitivity of the technology. While we have demonstrated that the technology can be used for the sequencing of simple and complex mixtures of hybridoma cells, we have not yet quantified the sensitivity and specificity of our antibody sequencing technology. This could be done by labelling different hybridoma cells with hashtag antibodies carrying specific barcodes (Stoeckius et al., 2018), a technology which is relatively well-established in the lab. If we take hashtag labelling as the gold standard, we could then quantitatively estimate the rates of antibody detection, relative to the hashtag labels. These could also be combined with spike-in experiments, where a hybridoma cell line expressing a known antibody could be spiked into a complex mixture of hybridoma cells at different ratios (1:100, 1:1,000, 1:10,000 etc) to provide a measure of the sensitivity and specificity of our technique.

The fact that the technology simply requires an additional emulsion PCR step

with appropriate primers suggests that it can easily adapted for the sequencing of antibody sequences of other species or for the sequencing of other variable mRNAs present within the transcriptome. While the data presented in this thesis have utilised the technology exclusively for the sequencing of mouse antibody heavy and light chains, the technology has been used in the lab for the sequencing and characterisation of human antibody heavy and light chains (Hongxing Hu, unpublished data), demonstrating its adaptability for use in different organisms. In addition, it is not difficult to envision its adaptation for the sequencing of T-cell receptors (TCRs), which resemble antibodies and B-cell receptors in terms of protein structures and the genetic mechanisms by which variability is produced.

There exist other emulsion-based antibody sequencing technologies that are compatible with simultaneous whole-transcriptome analysis (as previously discussed in Section 1.8.5), such as RAGE-seq (Singh et al., 2019), which uses long-read sequencing to sequence antibody sequences and their associated barcodes, as well as any other techniques that utilise additional oligonucleotide sequences to capture sequences of interest in a targeted manner, such as InDrop (Klein et al., 2015) or DART-seq (Saikia et al., 2019). However, these technologies frequently require an additional sequencing technique, namely nanopore sequencing in the case of RAGE-seq, or require prior manipulation and alteration of beads, which can be problematic to optimise.

In contrast, our technology does not require any prior manipulation of beads or additional sequencing techniques, which are both costly and difficult to optimise. Instead, it simply requires the addition of an emulsion PCR step after cDNA generation, which involves a round of droplet generation that is extremely similar to what is required in the Drop-seq protocol, and an additional round of PCR. The only additional reagents required are new primers and a DNA polymerase, which are relatively inexpensive, making our technology an affordable and simple option for the sequencing of antibody sequences, especially within labs that are familiar with Drop-seq.

However, our technology is currently not compatible with the 10X Genomics Single Cell 3' Gene Expression workflow. This is unfortunate, as the 10X Genomics workflows have been shown to perform better than Drop-seq and InDrop (Zhang et al., 2019). As the primer sequences utilised in the 10X pipelines are generally accessible, it is possible to adapt and optimise the technology for use with the 10X workflows in the future.

The ability to sequence antibody heavy and light chains, and to associate the chains with relevant transcriptomes, lays the groundwork for the development of a method that enables the study of the effect of antibodies on the transcriptomes of single target cells, permitting high-throughput target-agnostic functional antibody

screening. This technology will be outlined and further explained in the general discussion in Chapter 12.

12 General discussion and perspectives

Antibody screening continues to remain highly relevant in modern medicine (Urquhart, 2020a), but current established screening methods are either limited in their ability to detect functional antibodies (phage display) or are time-consuming, reagent-intensive and expensive, thus limiting the diversities of libraries that can be screened (hybridoma technology). These methods are also target- and function-specific, in contrast to our target-agnostic approach, which will be outlined below.

Droplet microfluidics has been employed to overcome various limitations of phage display and hybridoma technology, with the low sample volumes and high-throughputs permitting high-throughput functional antibody screening for antibodies against soluble targets (El Debs et al., 2012). In addition, while the screening for antibody binders has been carried out against whole cells (Shembekar et al., 2018), these technologies have not yet been demonstrated for the screening for functional antibodies against membrane-based targets or whole cells, and currently require prior knowledge about the target, such that screening is not target-agnostic.

We believe that droplet microfluidics continues to fulfil many criteria for the advancement of antibody screening. In particular, we believe that the use of droplet microfluidics for single-cell transcriptomics (as outlined in Section 1.6.4) can complement current droplet microfluidic antibody screening, to enable the development of a powerful method that permits high-throughput target- and function-agnostic functional antibody screening, by studying the effect of antibodies on single-cell transcriptomes.

12.1 Outline of the technology

The technology would be built upon the conclusions derived from Part I and the antibody sequencing technology developed in Part II.

Having successfully demonstrated that picoinjection can be utilised for the selective processing of droplets with desired occupancies (Section 4.3), we would

implement a workflow extremely similar to that outlined in Chapter 4. Firstly, we would co-encapsulate target cells with antibody-secreting cells. We would then incubate the droplets to permit antibody accumulation before carrying out picoinjection to add lysis buffer, ideally specifically to the droplets of interest, before reverse transcription and PCR are carried out (Fig. 12.1, (1) - (3)).

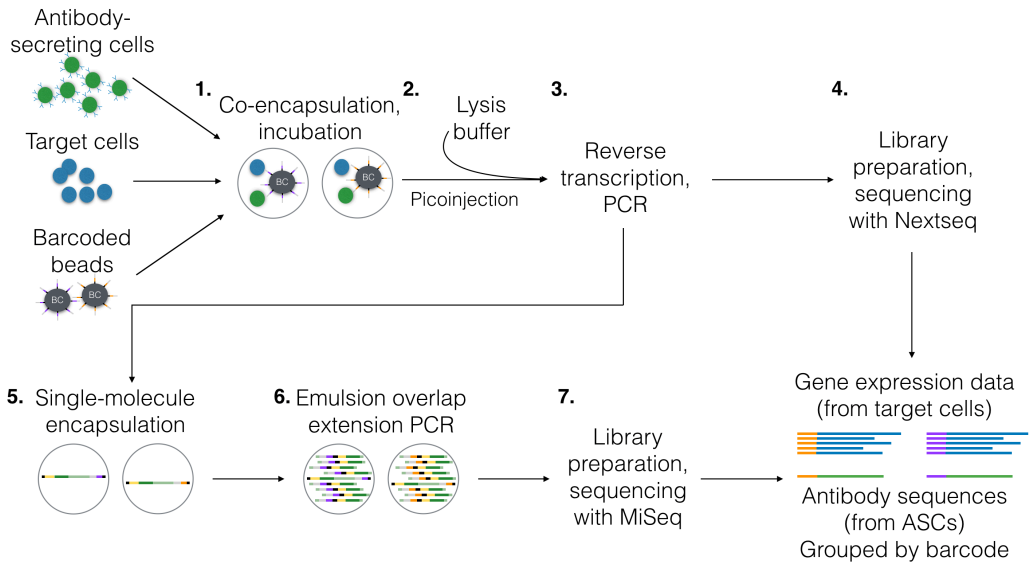


Figure 12.1: Microfluidic functional antibody screening utilising single-cell transcriptomics. (1) Antibody-secreting cells are co-encapsulated with target cells and incubated to enable the accumulation of antibodies and the induction of transcriptomic changes in the target cells. **(2)** Lysis buffer is then added via selective picoinjection, before **(3)** the emulsion is broken and reverse transcription and PCR are carried out. **(4)** Library preparation is carried out and the gene expression libraries are sequenced via NextSeq. The cell barcodes present enable all cDNA to be mapped back to their droplets of origin. **(5)** Concurrently, PCR products from (3) are encapsulated at a limiting dilution, such that a large majority of droplets contain either no, or a single PCR product. DNA polymerase and four different groups of primers are also included in the droplets, which facilitate the emulsion overlap extension PCR to produce a fragment compatible with sequencing. **(6)** Emulsion overlap extension PCR is carried out by thermocycling of the generated droplets. **(7)** Library preparation is carried out and the resulting fragments are sequenced on a MiSeq, giving the antibody heavy and light chain sequences from the antibody-secreting cells (ASCs). The cell barcodes (purple, orange) present enable the association of antibody heavy and light chain sequences with each other (green), and with cDNA of the associated target cell (blue).

The PCR products would then be utilised for two parallel applications. Firstly, they could be utilised for library preparation and sequencing, to yield the gene expression profiles of all cells, although we would be specifically interested in the gene expression profile of target cells (Fig. 12.1, (4)). Secondly, as outlined in Chapter 8 and validated in Chapters 9 and 10, the PCR products could be co-encapsulated with DNA polymerase and PCR primers. This permits the selective amplification of antibody heavy and light chain sequences, via an emulsion overlap extension PCR that fuses the associated cell barcodes and UMIs with the antibody sequences of interest into shorter fragments compatible with Illumina sequencing.

This allows the sequencing of antibody heavy and light chain sequences while ensuring that they could be associated to each other (Fig. 12.1, (5) - (7)).

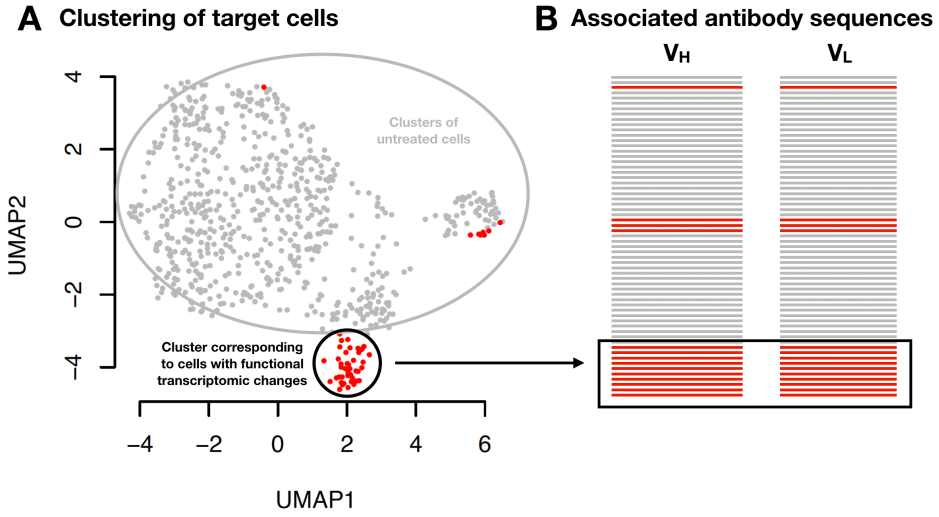


Figure 12.2: Data analysis schematic for microfluidic functional antibody screening utilising single-cell transcriptomics. Our sequencing data includes gene expression data from target cells **(A)** and antibody heavy and light chain variable sequences from the antibody-secreting cells **(B)**. **A.** Normalised gene expression data of 683 MCF-7 cells treated with DMF (negative control) (grey) and 50 MCF-7 cells treated with niclosamide (red), visualised using Uniform Manifold Approximation and Projection (UMAP), with dimensions = 20 for dimensionality reduction. This data was obtained during subsampling analysis and has been reproduced here with the permission of Xiaoli Ma, EPFL. This UMAP is utilised as a schematic, to illustrate that the gene expression data obtained from our future screens could be clustered to separate a background of untreated cells or cells which have no transcriptomic changes (grey) from clusters corresponding to cells that have undergone functional transcriptomic changes (red). **B.** Our sequencing data includes antibody heavy (V_H) and light chain variable sequences (V_L) from the antibody-secreting cells. Based on the target cells of interest identified from clustering, the antibody heavy and light chain sequences from hybridoma cells in the same droplets as those target cells can be identified via their common barcode sequences (black box). These antibody heavy and light chain sequences can then be used for cloning, synthesis and characterisation of these antibodies for further validation.

We would obtain two main categories of data, namely the gene expression data, of which we would be specifically interested in the data from the target cells (Fig. 12.2A), and the antibody heavy and light chain variable sequences from the antibody-secreting cells (Fig. 12.2B). Cell transcriptomes containing functional transcriptomic changes of interest can be identified (Fig. 12.2A, red cluster circled in black), with the concepts and ideas discussed in Chapters 6 and 7 or with the use of prior sequencing data for the same model system. The gene expression data from the target cells and the antibody heavy and light chain variable sequences can be grouped by barcode, such that the transcriptome of an individual target cell can be linked to the antibody sequences of their associated antibody-secreting cell for each droplet. This enables the antibody heavy and light chain variable regions of

interest to be identified (Fig. 12.2B, black box), and synthesised for further study and validation.

12.2 Characteristics of the technology

The technology has various advantages over current antibody sequencing techniques, which were outlined in Section 1.9.

Our technology analyses single antibody-secreting cells. This means that primary B-lymphocytes could be used for our screen, and that single hybridoma cells can be processed without the need for hybridoma cell expansion, which is both time-consuming and expensive, and limits the number of hybridoma clones that can be studied. The throughput of this technology would instead be largely restricted by sequencing and the associated costs, which will be elaborated on more in Section 12.3.

As no prior knowledge of specific protein targets or pathways is required, and the phenotypes of interest do not have to be defined beforehand, this antibody screening technique would be fully target- and function-agnostic. We expect that any antibody that changes the target cell transcriptome in any measurable way will be picked up in our screen. In addition, the use of target cell transcriptomes as a readout would ensure that any antibodies identified would be functional, rather than merely binding.

Overall, the technology is not unduly complex to implement. The single technologically-challenging step would be that of picoinjection, which requires lasers and their respective PMTs for the detection of cells within droplets, along with the associated LabVIEW software to measure peak fluorescence intensities of droplets and to gate the droplet populations of interest. The workstation also requires an electronic setup, to permit activation of an electric field on chip, to destabilise the surfactant film surrounding droplets and permit injection of reagents into droplets. In addition, the facilities for the manufacture of microfluidic chips are required, both for the production of PDMS chips, and the addition of low melting solder into the inlets for the generation of on-chip electrodes.

The rest of the protocol is not technologically more complex than conventional Drop-seq. The workflow for the sequencing of antibody heavy and light chain variable regions (as described in Chapter 8) does not require any prior manipulation of beads, but is simply viable with the addition of an emulsion overlap extension PCR to the Drop-seq protocol. The droplet generation carried out here is essentially analogous to the initial droplet generation in Drop-seq, and simply requires the addition of PCR reagents and primers.

However, our technology is not compatible with the 10X Genomics Gene Expression workflow, as the need for pre-incubation of the two cell types in the

absence of the lysis buffer is not possible with 10X Genomics proprietary reagents. As the 10X Genomics workflows have been shown to perform better than Drop-seq (Zhang et al., 2019), this is not ideal, but we believe that Drop-seq remains more than sufficient for our prospective antibody sequencing technology.

12.3 Remaining challenges

As mentioned above, our technology will largely be limited by sequencing costs. To mitigate this, we would need to optimise the identification of perturbations from a small number of cells and minimise the processing of droplets that do not provide useful data for our analyses, while exploring modifications such as targeted sequencing. These improvements seek to increase the throughput of our technology without a corresponding increase in costs. These correspond closely to the challenges that were outlined in Section 7.2, namely:

1. Elucidating and reducing the minimum cell number to pick up perturbations
2. Overcoming Poisson distribution during cell and bead encapsulation
3. Establishing targeted sequencing

As previously mentioned in Section 7.2, the identification of perturbations from a small number of cells is crucial in the functioning of our screening technology. Our analysis of a drug perturbation on two different cell lines suggests that between 30 - 100 cells would be required to detect a perturbation via clustering analysis (Chapter 6). While the effect size of a modulation caused by antibody binding may differ from that of drugs, we believe that this is a realistic estimate, especially as it is in line with similar analyses involving single-cell CRISPR screens with transcriptomic readouts (Alda-Catalinas et al., 2020; Datlinger et al., 2017; Dixit et al., 2016; Schraivogel et al., 2020).

This would be particularly problematic for the implementation of the technology for antibody screening, as we would like to screen a diverse library of antibody-secreting cells, while the requirement for 30 - 100 single cells per antibody clone would permit us to only realistically screen hundreds, or maximally, thousands of different antibody clones, before sequencing costs become exorbitant.

In addition, as previously mentioned in Section 7.2, this would require a population of antibody-secreting cells where genetically identical antibody secreting cells are present, with at least 30 individual cells per clone. This is realistic for hybridoma cells, as they can be easily cultured and expanded, but would be more difficult to achieve for primary plasma B-lymphocytes, which are extremely diverse and are difficult to culture. This suggests that the reduction of the number of cells required to detect a perturbation is an area that requires improvement.

We believe that clustering may not necessarily be the best bioinformatic tool to detect perturbations. Instead, alternative strategies should be investigated. For example, it is not unimaginable that the examination of a subset of highly differentially expressed and biologically relevant genes could enable a more effective segregation of treated and untreated populations, even in the absence of targeted sequencing. Similarly, various pipelines have been, and continue to be developed for the analysis of transcriptomic data obtained from single-cell CRISPR screens. These pipelines, such as MOFA+ (Alda-Catalinas et al., 2020; Argelaguet et al., 2020) and MUSIC (Duan et al., 2019) could be easily adapted for use in our screening pipeline, but their efficacy remains to be evaluated.

However, it must be noted that any screen for biologically relevant antibody-secreting cells would be slightly different from single-cell CRISPR screens with transcriptomic readouts. Those screens typically examine one, or a few, specific biological pathways, while our screen aims to be target-agnostic and intends to detect hits from a plethora of different pathways. In addition, while these single-cell CRISPR screens value both sensitivity and specificity very highly, our screen could accept slightly lower sensitivity (i.e. higher numbers of false negatives), as we would simply like to detect some biologically interesting perturbations, without necessarily having to reliably detect all relevant perturbations.

In the context of high-throughput antibody screening, it would be extremely vital to carry out both a subsampling analysis (as outlined in Chapter 6) and a model screen, with a model system involving an antibody secreting cell line, together with the relevant target cell line. While the use of an additional model system would expand the conclusions that we have made from our currently limited dataset, characterisation of a dataset involving a model system with antibody secretion would ensure that we are well-placed for further investigation of more complex populations of antibody-secreting cells, and for the optimisation of the technology for antibody screening.

The regulation of droplet occupancy also remains a challenge for high-throughput target-agnostic antibody screening via transcriptomics. We would like to maximise the number of droplets containing two cells, namely one cell of each cell type, but this is challenging to achieve with Poisson distribution, as explained in Chapter 4. We are currently capable of selectively processing droplets containing one cell type, by selectively picoinjecting lysis buffer only into droplets in which a specific cell type is detected, as outlined in Section 4.3. Improving upon picoinjection to enable detection of multiple fluorescence signals and thus decision-making based on multiple cell types, will improve the accuracy of selective processing and minimise the sequencing of cells from undesirable droplets. This is not unimaginable, but will have to be established and optimised.

Alternatively, a higher level of regulation of encapsulation would enable the

generation of a higher percentage of droplets with the desired occupancy. This would enhance the efficiency of our antibody screening technology while reducing the unnecessary sequencing of uninteresting droplets. There are various possibilities on how this can be done, including the utilisation of deterministic cell encapsulation with Dean's force (Kemna et al., 2012; Schoeman et al., 2014) or hydrodynamic traps (Sauzade and Brouzes, 2017), but it remains to be seen how effective these methods may be.

Targeted sequencing will similarly be crucial in reducing sequencing costs, and could enable an increase in screening throughput without a corresponding increase in sequencing costs. The different techniques all aim to selectively enrich for specific sequences of interest, and this can be primarily done at three different steps of the emulsion-based single-cell RNA-seq protocol, namely during RNA capture, PCR amplification and enrichment via hybridisation capture, as explained in Section 7.2.

However, targeted sequencing requires a pre-selection of genes to be observed. While a hypothesis-driven selection of genes may actually make it easier to identify cells of interest (Schraivogel et al., 2020), this information may not be available for all processes, and makes any screen no longer fully target-agnostic.

It is clear that while our technology is very promising, it continues to be impaired by the high sequencing costs involved and by various challenges associated with these high sequencing costs, which limit the screening throughput. This suggests that improvements in the three areas mentioned, namely reducing the minimum cell number to pick up perturbations, overcoming Poisson distribution during encapsulation, and establishing targeted sequencing, are crucial in increasing the viability and desirability of the technology. Nonetheless, given that sequencing costs are rapidly decreasing over time (Hall, 2013; *The Cost of Sequencing a Human Genome*), we expect that the technology will be increasingly more viable and more desirable in the future.

12.4 Outlook

We believe that this screening technology can be easily adapted for the study and screening of interactions of T-cells with their interaction partners, including but not limited to, antigen-presenting cells (APCs) and dendritic cells. This is as the protein structures of and the genetic mechanisms by which variability is produced in T-cell receptors are highly similar to that of antibodies and B-cell receptors.

Droplet microfluidics has greatly revolutionised both the field of single-cell sequencing and the field of antibody screening. We envision that a future technology combining both applications of microfluidics will permit high-throughput target-agnostic functional antibody screening and enable the identification of novel, biologically interesting antibodies.

In this thesis, we have assessed the feasibility of such a technology. Firstly, we have overcome the problem of Poisson distribution in droplet generation, where we successfully demonstrated that picoinjection can be utilised for the selective processing of droplets with desired occupancies. We have also identified two model systems suitable for the further optimisation of the single-cell transcriptomic analysis pipeline, namely the CHO-GLP1R-GFP and HEK293T-mCherry-Exendin-4 model system, where the cells interact via direct cell contact, and the system involving the 60bca hybridoma cell line and CD14-expressing U937 cells, where the interaction takes place indirectly, via the action of antibodies. Work has also been done to investigate the identification of a small number of positive events from a larger background of negative events, allowing us to frame further strategies to reduce the number of positive events needed to identify a perturbation.

Additionally, we have developed an antibody sequencing pipeline compatible with Drop-seq, and demonstrated its use on a predefined mixture of hybridoma cells and on a complex mixture of hybridoma cells for which we had no prior knowledge of antibody sequences.

Our work in these two areas has laid the groundwork for the development and optimisation of a technology utilising both these aspects for high-throughput target-agnostic functional antibody screening. While challenges remain, we believe that this antibody screening technology will be viable in the near future.

13 Materials and Methods

13.1 Methods in microfluidics

13.1.1 Mask design and production

The microfluidic devices utilised as part of this thesis work were first produced as photomasks and subjected to photolithography. They were designed in AutoCAD (Autodesk Inc., USA) and were either adapted from previous designs found in the lab or from publications. Designs were printed (Selba S.A. Versiox Switzerland) as negative masks (clear designs on a dark barkground) on transparent plastic sheets at a resolution of 25,400 dpi.

13.1.2 Photolithography for production of moulds

Based on the size of the design to be produced, silicon wafers (Siltronix, Silicon Materials, Germany) of an appropriate size ($3''$ or $4''$) were selected and cleaned with an air gun and heated on a hot plate, at 140 °C for 10 min. The wafers were then coated with a layer of SU-8 photoresist (MicroChem Corp., Newton MA). Different variants of SU-8 photoresist exist, which differ in viscosity, and thus in the height of the layers that can be formed. For example, the SU-8 2050 photoresist was used to attain channel depths ranging from 50 – 100 µm. Approximately 5 ml of the resist was dispensed on the wafer and spread uniformly on the wafer by a spin coater (WS-400BZ-6NPP/LITE, Laurell Technologies Corporation, USA). The time and spinning speed used were based on the manufacturer's recommendations. The wafers were then baked at 65 °C and 95 °C based on the manufacturer's instructions and the desired channel heights. The photomask was then placed on the wafer and exposed to UV light for a defined time (based on manufacturer's instructions) in a mask aligner (Karl Suss MA45). Crosslinking occurs at the sections exposed to UV light, such that the design on the photomask will be transferred to the resist. The wafers were once again baked at 65 °C and 95 °C based on the manufacturer's instructions and on the desired channel heights, before being developed with mr-Dev-600 (micro resist technology GmbH, Berlin), which removes all non-polymerised photoresist, leaving the desired structures behind. Finally, a

final baking step was carried out at 150 °C to solidify the structures on the mold. The height of the structures obtained was verified using a Profilometer (Faulhaber).

13.1.3 Manufacturing of microfluidic chips

Polydimethylsiloxane (PDMS) and curing agent (Sylgard 184 silocone elastomer kit, Dow Corning Corp, USA) were mixed at a 9:1 (w/w) ratio and degassed in a dessicator until all bubbles were removed. The mixture was then poured over moulds placed in petri dishes and baked overnight at 65 °C. The solidified PDMS was then cut with a scalpel and peeled off the mould. Inlets for fluid and for electrodes were punched with biopsy punches with a diameter of 0.75 mm (World Precision Instruments, USA). The PDMS was then cleaned using an air gun and with adhesive tape, before being plasma bonded (Femto, Diener electronic GmbH + Co. KG, Germany) to untreated glass slides or to the non-conductive sides of indium tin oxide (ITO) glass slides (chips for picoinjection) (CG 41IN-S207, Delta Technologies Limited, USA). Picoinjection devices were also placed on a hotplate at 95 °C until the PDMS was heated up, before low melting solder was inserted into the inlets for the electrodes. These were then connected to cables at the inlets, which could then be connected to the electronic setup of the workstation. To make channel surfaces hydrophobic, Aquapel (Autoserv, Germany) was flushed through the channels and subsequently removed by flushing with air, before being left to dry for 10 min on a hot plate set to 80 °C.

13.1.4 Injection of fluids into microfluidic chips

Luer-Lok syringes (Beckon-Dickinson, USA) of an appropriate size were loaded with the reagents to be injected, before being connected to 25G 1″ needles (Beckon-Dickinson, USA), which were in turn connected to PTFE tubing with an inner diameter of 0.32 mm and an outer diameter of 0.78 mm (TW30, Adtech Polymer Engineering Ltd, UK). Syringes were primed to remove all air bubbles and fill the tubing with reagent. These syringes were then mounted on PHD 22/2000 Syringe Pumps (Harvard Apparatus, USA), and the syringe diameters and desired flow rates were programmed. The free ends of each tubing were connected to their respective tubing inlet in the microfluidic chip, ensuring that reagents were injected into the chip at defined flow rates.

13.1.5 Optical setup for imaging and fluorescence measurements

A standard inverted light microscope (Eclipse Ti-S, Nikon GmbH, Germany) was used for bright-field imaging inside droplets. A high speed camera (Motion BLITZ EoSens mini1, Mikrotron GmbH, Germany) was used to make high-speed recordings during droplet generation.

Spectroscopic fluorescence measurements were carried out with an in-house optical system, which is comprised of a vibration-reduced breadboard fitted with a

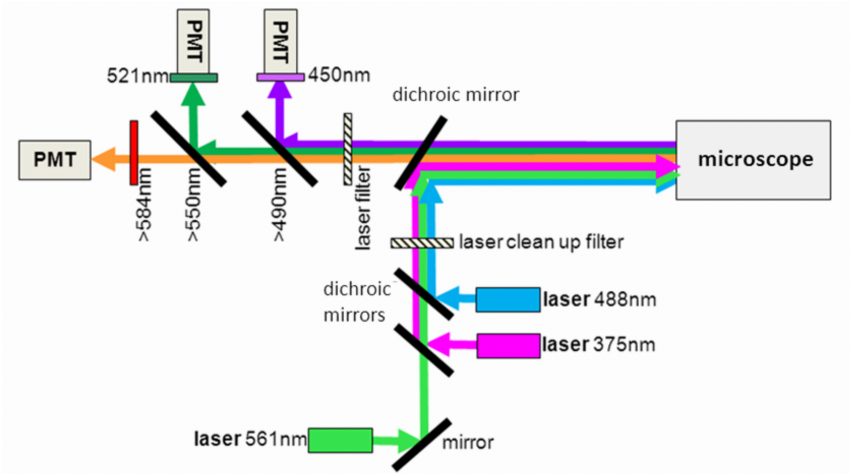


Figure 13.1: Optical setup of the microfluidic workstation. Three different lasers are utilised for excitation at wavelengths of 375 nm, 488 nm and 561 nm. Emission is measured with three PMTs, at wavelengths of 450 nm, 521 nm and >584 nm. Figure reprinted from Frenzel and Merten, 2017 under a Creative Commons Attribution License (CC-BY 3.0) - Published by The Royal Society of Chemistry.

standard inverted microscope, as mentioned above. Lasers of wavelengths 375 nm, 488 nm and 561 nm were directed over dichroic mirrors into the microscope objective (Fig 13.1). The emitted light was then directed through bandpass filters to a set of three photomultiplier tubes (PMTs) (Hamamatsu), enabling the simultaneous excitation and detection of three different parameters (Fig 13.1). For all measurements, a 40x objective was used. In the experiments outlined in this thesis, only the lasers of wavelengths 375 nm and 488 nm and their respective PMTs were utilised.

13.1.6 Selective picoinjection

Cell lysis reagents were injected into the picoinjector inlet at 200 µl/h. The droplet emulsion was mounted onto a syringe pump which was set to inject at 500 µl/h. Once the air was removed from the tubing, the tubing was connected to the chip and the injection rate lowered to 100 µL/h. Injected droplets were spaced out with oil, and the flow rates of the droplets and the oil were adjusted to achieve good spacing and a constant flow of droplets.

We used an in-house LabVIEW software (developed by Ramesh Utharala) to measure peak fluorescence intensities of droplets and to set gates for the droplet population of interest. When a droplet of interest passes the laser, an electrical pulse is applied to induce the addition of cell lysis reagents in the nozzle into the droplet of interest.

The picoinjection technology was kindly shared with us by Michael Ryckelynck from the University of Strasbourg.

13.1.7 Cell and bead encapsulation for Drop-seq and picoinjection

Cells to be used for picoinjection were washed in PBS before being stained with CellTrace™ CFSE (Thermo Fisher Scientific Inc., USA) as per manufacturer's instructions. Alternatively, cells for Drop-seq were unstained. Cells were then washed twice with PBS before being counted and resuspended in PBS with 0.01% Bovine Serum Albumin (BSA) (Sigma-Aldrich, USA) at a concentration of 4 x 10^6 cells/ml before being diluted to a concentration corresponding to a final $\lambda = 0.05$ after encapsulation.

Approximately 120,000 Drop-seq barcoded beads (ChemGenes, USA) (Macosko et al., 2015) were spun down at 1000 x *g* at 4 °C and resuspended in 950 µl lysis buffer and 50 µl 1M dithiothreitol (DTT) (Thermo Fisher Scientific Inc., USA).

Cells and beads were then loaded into separate 3 ml Luer-Lok syringes (Beckon-Dickinson, USA) directly connected to TW30 tubing (Adtech Polymer Engineering Ltd, UK) via PDMS plugs. A magnetic mixing disc (VP Scientific, USA) was also included and rotated at a low speed using a magnetic mixer (Combimag Reo, IKA, Germany), to ensure even cell mixing. The syringes containing cells or beads were mounted on syringe pumps with the syringes pointing downwards. The chip design used was the same as that of the devices used for Drop-seq (Macosko et al., 2015), with a height of 80 µm.

The continuous phase used was 1% PicoSurf-1 (Sphere Fluidics, UK) in HFE-7500 Novec Engineered Fluid (3M, USA) for experiments involving picoinjection. For Drop-seq experiments without picoinjection, QX200™ Droplet Generation Oil for EvaGreen (Bio-Rad Laboratories Inc., USA) was utilised. The oil phase was injected at a flow rate of 12,000 µL/h, while the cells and beads were injected at flow rates of 3000 µL/h.

13.1.8 Co-encapsulation of CHO-GLP1R-GFP cells with HEK293T-mCherry-Exendin-4 cells

CHO-GLP1R-GFP cells were stained with CellTrace™ Violet (Thermo Fisher Scientific Inc., USA) as per manufacturer's instructions. CHO-GLP1R-GFP cells, HEK293T-WT and HEK293T-mCherry-Exendin-4 cells were washed once with PBS before being counted and resuspended in FreeStyle 293 Expression Media (Gibco, Thermo Fisher Scientific Inc., USA) at a concentration corresponding to a final $\lambda = 1$ for each cell type after encapsulation.

Cells were then loaded into separate 3 ml Luer-Lok syringes (Beckon-Dickinson, USA) directly connected to TW30 tubing (Adtech Polymer Engineering Ltd, UK) via PDMS plugs. A magnetic mixing disc (VP Scientific, USA) was also included and rotated at a low speed using a magnetic mixer (Combimag Reo, IKA, Germany), to ensure even cell mixing. The syringes were mounted on syringe pumps with the

syringes pointing downwards. The chip design used was the same as that of the devices used for Drop-seq (Macosko et al., 2015), with a height of 80 µm.

The continuous phase used was 1% PicoSurf-1 (Sphere Fluidics, UK) in HFE-7500 Novec Engineered Fluid (3M, USA), and was injected at a flow rate of 12,000 µL/h, while the cells were injected at flow rates of 3000 µL/h. Two samples were prepared, where CHO-GLP1R-GFP cells were co-encapsulated together with either HEK293T-WT or HEK293T-mCherry-Exendin-4 cells.

Droplets were incubated for 10 h at 37 °C, 5% CO_2 and in a humidified environment, prior to emulsion breakage with 0.3x volume of 1H,1H,2H,2H-Perfluorooctanol (PFO) (ABCR GmbH, Germany) and analysis with flow cytometry.

13.1.9 Droplet generation for single-molecule PCR

2X emulsion linkage PCR mix was prepared as outlined in Section 13.14.2 and DNA obtained after cDNA purification was diluted to 3 x 10^{-4} ng/µl with water containing 0.5 mg/ml Ultrapure BSA (Thermo Fisher Scientific Inc., USA). Both mixtures were loaded in 3 ml Luer-Lok syringes (Beckon-Dickinson, USA) directly connected to TW30 tubing (Adtech Polymer Engineering Ltd, UK) via PDMS plugs. The syringes were mounted on syringe pumps with the syringes pointing downwards.

The chip design used was the same as the design utilised for the PDMS co-flow droplet generation design used in conventional Drop-seq (Macosko et al., 2015), but with a height of 55 µm, to generate droplets with 70 µm diameter and a volume of about 180 pl.

The continuous phase used was QX200™ Droplet Generation Oil for EvaGreen (Bio-Rad Laboratories Inc., USA), which was injected at a flow rate of 12,000 µL/h, while the PCR mix and DNA templates were injected at flow rates of 1800 - 2000 µL/h.

13.2 Mammalian cell culture

13.2.1 Cell culture conditions

SK-BR-3 cells were obtained from the Jechlinger lab, EMBL Heidelberg, and were cultured in DMEM High Glucose media (Gibco, Thermo Fisher Scientific Inc., USA) supplemented with 10% fetal bovine serum (FBS) (Gibco, Thermo Fisher Scientific Inc., USA), 1 mm sodium pyruvate (Gibco, Thermo Fisher Scientific Inc., USA), 100 units/ml Penicillin and 100 µg/ml Streptomycin (Gibco, Thermo Fisher Scientific Inc., USA). The cells were subcultured 2 - 3 times a week at a ratio of 1:5.

K-562 cells (ATCC®CCL-243™, USA), and were cultured in IMDM media (Gibco, Thermo Fisher Scientific Inc., USA) supplemented with 10% FBS (Gibco,

Thermo Fisher Scientific Inc., USA), 100 units/ml Penicillin and 100 µg/ml Streptomycin (Gibco, Thermo Fisher Scientific Inc., USA). The cells were subcultured 3 times a week at a ratio of 1:10 - 1:15.

Hybridoma cell lines were cultured in either DMEM High Glucose (4D5, OKT-9, OKT-3, anti-CD4, H25B10) (Gibco, Thermo Fisher Scientific Inc., USA) or RPMI-1640 media (60bca, 8G5F11, IE9F9, H25B10) (Gibco, Thermo Fisher Scientific Inc., USA) supplemented with 10% FBS (Gibco, Thermo Fisher Scientific Inc., USA) and 100 units/ml Penicillin and 100 µg/ml Streptomycin (Gibco, Thermo Fisher Scientific Inc., USA). The cells were subcultured 2 - 3 times a week at a ratio of 1:6 – 1:20.

Table 13.1: Hybridoma cell lines utilised

Hybridoma cell line	**Company**	**Catalog number**
4D5	ATCC, USA	CRL-10463
OKT-9	ATCC, USA	CRL-8021
OKT-3	ATCC, USA	CRL-8001
H25B10	ATCC, USA	CRL-8017
60bca	ATCC, USA	HB-247
8G5F11	Kerafast, USA	EB0010
IE9F9	Kerafast, USA	EB0012
Precision Antibody hybridoma (against SK-BR-3 cells)	Precision Antibody, USA	NA

Antibody sequences of the anti-CD4 hybridoma cell line can be found in Appendix G.

Precision Antibody hybridoma cells were cultured in ClonaCell^{TM}-HY Medium C (STEMCELL Technologies Inc., Canada), supplemented with HAT Supplement (sodium hypoxanthine (0.1 mm), aminopterin (0.4 µM) and thymidine (16 µM)) (Gibco, Thermo Fisher Scientific Inc., USA) and 100 units/ml Penicillin and 100 µg/ml Streptomycin (Gibco, Thermo Fisher Scientific Inc., USA). Information on the generation of the Precision Antibody hybridomas can be found in Chapter 3 and in Subsection 13.13.2.

U937 cells (ATCC®CRL-1593.2^{TM}, USA) were cultured in RPMI-1640 media (Gibco, Thermo Fisher Scientific Inc., USA) supplemented with 10% FBS (Gibco, Thermo Fisher Scientific Inc., USA), 100 units/mL Penicillin and 100 µg/mL Streptomycin (Gibco, Thermo Fisher Scientific Inc., USA). The cells were subcultured 2 – 3 times a week at a ratio of 1:5.

Wild-type Chinese Hamster Ovary (CHO-K1) cells were obtained from the Ellenberg lab, EMBL Heidelberg, while the CHO-GLP1R-GFP cells were obtained from the McDonald lab at the Scripps Research Institute (Zhang et al., 2015). They were cultured in F12-K media (Gibco, Thermo Fisher Scientific Inc., USA) supplemented with 10% FBS (Gibco, Thermo Fisher Scientific Inc., USA), 2 mm L-Glutamine (Sigma-Aldrich, USA), 100 units/ml Penicillin and 100 µg/ml

Streptomycin (Gibco, Thermo Fisher Scientific Inc., USA). The cells were subcultured 2 – 3 times a week at a ratio of 1:10 – 1:15, using 0.05% Trypsin-EDTA (Gibco, Thermo Fisher Scientific Inc., USA) to detach the cells from the bottom of the flasks.

The HEK293T cells expressing mCherry and Exendin-4 were generated in-house by Amanda Vanselow, using a plasmid obtained from the McDonald lab at the Scripps Research Institute (Zhang et al., 2015). HEK293T cells (wild-type (WT) (ATCC®CRL-3216™, USA) and mCherry-Exendin-4) were cultured in DMEM High Glucose media (Gibco, Thermo Fisher Scientific Inc., USA) supplemented with 10% FBS (Gibco, Thermo Fisher Scientific Inc., USA), 2 mm L-Glutamine (Sigma-Aldrich, USA), 1 mm sodium pyruvate (Gibco, Thermo Fisher Scientific Inc., USA), 100 units/ml Penicillin and 100 µg/ml Streptomycin (Gibco, Thermo Fisher Scientific Inc., USA). The cells were subcultured 3 times a week at a ratio of 1:4 – 1:6, using 0.05% Trypsin-EDTA (Gibco, Thermo Fisher Scientific Inc., USA) to detach the cells from the bottom of the flasks.

MCF-7 cells were obtained from the Jechlinger lab, EMBL Heidelberg. They were cultured in DMEM High Glucose media (Gibco, Thermo Fisher Scientific Inc., USA) supplemented with 10% FBS (Gibco, Thermo Fisher Scientific Inc., USA), 2 mm L-Glutamine (Sigma-Aldrich, USA), 1X non-essential amino acids (Sigma-Aldrich, USA), 100 units/ml Penicillin and 100 µg/ml Streptomycin (Gibco, Thermo Fisher Scientific Inc., USA). The cells were subcultured 2 – 3 times a week at a ratio of 1:6 - 1:10, using 0.05% Trypsin-EDTA (Gibco, Thermo Fisher Scientific Inc., USA) to detach the cells from the bottom of the flasks.

A375-P cells were obtained from the Diz-Muñoz lab, EMBL Heidelberg. They were cultured in RPMI-1640 media (Gibco, Thermo Fisher Scientific Inc., USA) supplemented with 10% FBS (Gibco, Thermo Fisher Scientific Inc., USA), 2 mm L-Glutamine (Sigma-Aldrich, USA), 100 units/ml Penicillin and 100 µg/ml Streptomycin (Gibco, Thermo Fisher Scientific Inc., USA). The cells were subcultured 2 – 3 times a week at a ratio of 1:6 - 1:10, using 0.05% Trypsin-EDTA (Gibco, Thermo Fisher Scientific Inc., USA) to detach the cells from the bottom of the flasks.

All cells were cultured at 37 °C, 5% CO_2 and in a humidified environment.

13.2.2 Obtaining hybridoma supernatants

Hybridoma cells were washed with PBS and seeded in their respective media (Table 13.2) and incubated at at 37 °C, 5% CO_2 and in a humidified environment for 24 hours. Hybridoma cells and supernatant were then spun at 500 x *g* for 5 min, before the supernatant was filtered through a 0.45 µm Millex-HV Filter, PVDF (Merck Millipore, MilliporeSigma, USA) and aliquoted and stored at -20 °C.

Table 13.2: Hybridoma supernatant media

Media composition	Hybridoma cell lines
DMEM High Glucose (Gibco, Thermo Fisher Scientific Inc., USA) supplemented with 10% FBS (Gibco, Thermo Fisher Scientific Inc., USA) and 100 units/ml Penicillin and 100 µg/ml Streptomycin (Gibco, Thermo Fisher Scientific Inc., USA)	4D5, OKT-9, OKT-3, anti-CD4, H25B10
RPMI-1640 (Gibco, Thermo Fisher Scientific Inc., USA) supplemented with 10% FBS (Gibco, Thermo Fisher Scientific Inc., USA) and 100 units/ml Penicillin and 100 µg/ml Streptomycin (Gibco, Thermo Fisher Scientific Inc., USA)	8G5F11, IE9F9, H25B10
Hybridoma Serum-free media (Gibco, Thermo Fisher Scientific Inc., USA) supplemented with 100 units/ml Penicillin and 100 µg/ml Streptomycin (Gibco, Thermo Fisher Scientific Inc., USA)	60bca, H25B10
ClonaCell^{TM}-HY Medium E (STEMCELL Technologies Inc., Canada) supplemented with 100 units/ml Penicillin and 100 µg/ml Streptomycin (Gibco, Thermo Fisher Scientific Inc., USA)	Precision Antibody hybridomas, H25B10

13.2.3 Lentivirus production

HEK293T cells were seeded and left to adhere overnight, to give approximately 60-70% confluency. 10 µg of pLVX-IRES-ZsGreen (transfer plasmid), 10 µg of pCMV-dR8.2 (packaging plasmid) and 5 µg of pMD2.G with VSV-G (envelope plasmid) were utilised for the transfection of each T175 flask of HEK293T cells, with calcium phosphate used as the transfection reagent, according to the manufacturer's instructions (Sigma-Aldrich, USA). The media was changed 24 h after transfection, and lentivirus supernatants were harvested at 48 h and 72 h after transfection and pooled. The media used for the transfection and collection of virus supernatants contain 5 mm sodium butyrate (Sigma-Aldrich, USA).

The supernatants were concentrated with Lenti-X^{TM} Concentrator (Takara Bio, Japan), according to the manufacturer's instructions. The lentivirus was resuspended in the appropriate media for the cell type to be transduced, with the addition of 5 mm sodium butyrate (Sigma-Aldrich, USA) and 5 µg/mL polybrene (Merck Millipore, MilliporeSigma, USA).

13.3 Cell treatment and activation

All hybridomas supernatants are produced and utilised as indicated in Table 13.2. Different hybridoma supernatants utilised in same experiment use the same media composition.

13.3.1 Treatment of SK-BR-3 cells with recombinant anti-Her-2 antibody

SK-BR-3 cells were unstimulated or treated with 1 µg/ml Anti-HER2-Tra-hIgG1 (InvivoGen, USA) for 6 h at 37 °C, 5% CO_2 and in a humidified environment, prior to cell harvesting and library generation.

13.3.2 Treatment of Jurkat cells with anti-CD4 hybridoma supernatant

Jurkat cells were washed with and resuspended in FreeStyle 293 Expression Media (Gibco, Thermo Fisher Scientific Inc., USA). Samples were mixed at a 1:1 ratio with anti-CD4 hybridoma supernatant or with media only (negative control) and incubated for 6 h at 37 °C, 5% CO_2 and in a humidified environment, prior to cell harvesting and library generation.

13.3.3 Treatment of Jurkat cells with anti-CD3 antibody and PMA

Jurkat cells were treated with 1 µg/ml anti-His (negative control) (Biolegend, USA) or 1 µg/ml anti-CD3 antibody (Biolegend, USA) for 8 h at 37 °C, 5% CO_2 and in a humidified environment, prior to cell harvesting and library generation. This took place concurrently with PMA treatment, if applicable.

Jurkat cells were incubated with 10 ng/ml of Phorbol 12-Myristate 13-Acetate (PMA) (InVivoGen, USA) for 8 h at 37 °C, 5% CO_2 and in a humidified environment, as outlined in Smeets et al., 2012.

13.3.4 Treatment of Jurkat cells with OKT-3 hybridoma supernatant

Jurkat cells were washed with and resuspended in FreeStyle 293 Expression Media (Gibco, Thermo Fisher Scientific Inc., USA). Samples were mixed at a 1:1 ratio with OKT-3 hybridoma supernatant or with H25B10 hybridoma supernatant (negative control) and incubated for 4 h at 37 °C, 5% CO_2 and in a humidified environment. Both hybridoma supernatants were used at a IgG concentration of 6.720 µg/mL. Concentrations were determined by ELISA (Appendix A).

13.3.5 Treatment of K-562 cells with OKT-9 hybridoma supernatant

For the bulk sequencing experiment, K-562 cells were treated with 1 µg/ml of anti-BSA antibody (negative control) (Thermo Fischer Scientific Inc., USA) or mixed at a 1:1 ratio with OKT-9 hybridoma supernatant (3.14 µg/ml, as determined by ELISA (Appendix A)) for 8 h at 37 °C, 5% CO_2 and in a humidified environment, prior to cell harvesting and library generation.

For the 10X sequencing experiment, K-562 cells were washed with and resuspended in FreeStyle 293 Expression Media (Gibco, Thermo Fisher Scientific Inc., USA). Samples were mixed at a 1:1 ratio with OKT-9 hybridoma supernatant or with H25B10 hybridoma supernatant (negative control) and incubated for 4 h

at 37 °C, 5% CO_2 and in a humidified environment. Both hybridoma supernatants were used at a IgG concentration of 3.14 µg/mL. Concentrations were determined by ELISA (Appendix A).

13.3.6 Treatment with niclosamide

MCF-7 and A375-P cells were treated with dimethylformamide (DMF) (Sigma-Aldrich, USA) (negative control) or 10 µM niclosamide (MedChemExpress LLC., USA) for 6 h at 37 °C, 5% CO_2 and in a humidified environment, prior to cell harvesting and further processing.

13.3.7 Differentiation of U937 with 1,25-dihydroxyvitamin D3 and treatment with anti-CD14 hybridoma supernatant

U937 cells were incubated with with 100 nm 1,25-dihydroxyvitamin D3 (Sigma-Aldrich, USA) for 24 h to stimulate CD14 expression (Baek et al., 2009). CD14-expressing U937 cells were washed with and resuspended in FreeStyle 293 Expression Media (Gibco, Thermo Fisher Scientific Inc., USA) containing 1 µg/mL lipopolysaccharides from *Salmonella minnesota* (Sigma-Aldrich, USA), 2% human serum (Sigma-Aldrich, USA), 100 units/ml Penicillin and 100 µg/ml Streptomycin (Gibco, Thermo Fisher Scientific Inc., USA)). These were mixed at a 1:1 ratio with 60bca hybridoma supernatant or with H25B10 hybridoma supernatant (negative control), and incubated for 6 h at 37 °C, 5% CO_2 and in a humidified environment. Both hybridoma supernatants were used at a IgG concentration of 5.37 µg/mL. Concentrations were determined by ELISA (Appendix A).

13.3.8 Treatment of CHO-GLP1R-GFP cells with Exendin-4

CHO-GLP1R-GFP cells were treated with 10 µM Exendin-4 (GenScript Biotech, USA) for 6 h at 37 °C, 5% CO_2 and in a humidified environment, prior to cell harvesting with 0.05% Trypsin-EDTA (Gibco, Thermo Fisher Scientific Inc., USA).

13.3.9 Co-cultivation of CHO-GLP1R-GFP cells with HEK293T-mCherry-Exendin-4 cells

HEK293T-WT or HEK293T-mCherry-Exendin-4 cells were seeded and permitted to adhere overnight. CHO-GLP1R-GFP cells were stained with CellTrace™ Violet (Thermo Fisher Scientific Inc., USA) as per manufacturer's instructions. CHO-GLP1R-GFP cells were then seeded and incubated for 0, 6, 10 or 30 h at 37 °C, 5% CO_2 and in a humidified environment, prior to cell harvesting with 0.05% Trypsin-EDTA (Gibco, Thermo Fisher Scientific Inc., USA).

13.3.10 Inhibition of lentiviral transduction by 8G5F11 and IE9F9 hybridoma supernatants

Lentivirus is added at a concentration that gives a final multiplicity of infection (MOI) of 5 (in media for K-562 culture with 10 mm sodium butyrate (Sigma-Aldrich, USA) and 10 µg/mL polybrene (Merck Millipore, MilliporeSigma, USA)), and mixed with an equal volume of 7.5 µg/mL H25B10 (negative control), 8G5F11 or IE9F9 hybridoma supernatant and incubated at 37 °C for 1 hour. The lentivirus-antibody mixture was incubated with K-562 cells overnight, with the media changed after 21 hours. K-562 cells were harvested for flow cytometry analysis 96 hours after viral transduction.

13.3.11 Cell cycle synchronisation with palbociclib

MCF-7 cells were treated with 2.5 µM palbociclib (Sellect Chemistry LLC., USA) for 24 h. The media containing palbociclib was removed and the cells were washed gently with PBS prior to subsequent analyses.

13.4 Bulk mRNA-sequencing

13.4.1 Library preparation and sequencing

Cells were washed twice with PBS, before RNA extraction was carried out with the RNeasy Mini Kit with on-column DNA digestion (as per the manufacturer's instructions) (QIAGEN N.V., Netherlands). For each cell type, treated and untreated samples were prepared, with three biological replicates prepared for each treatment condition.

The RNA obtained was analysed using the Bioanalyzer RNA 6000 Pico Assay and a 2100 Bioanalyzer (Agilent Technologies Inc., USA), before being used for library preparation via the Illumina TruSeq mRNA Stranded Library kit (Illumina Inc., USA), with the three replicates for each treatment condition being processed independently and the samples differentially indexed at the Adapter Ligation step. The libraries obtained were subject to fragment size analysis using High Sensitivity DNA Chips and a 2100 Bioanalyzer (Agilent Technologies Inc., USA). All samples were pooled at equimolar ratios and sequenced on a NextSeq 500 machine with a 75 cycles high output kit (Illumina Inc., USA), with 85 bases read for read 1 and 7 bases for the index reads.

13.4.2 Data analysis

The library was first demultiplexed using the in-line i7 Illumina indices to assign reads to samples. Demultiplexing was carried out by Jonathan Landry from the EMBL Genomic Core Facility.

An analysis pipeline was created and the analysis was performed in Galaxy. Fragments were trimmed with Trimmomatic (Galaxy Version 0.36.6) (Bolger et al., 2014) prior to mapping to the hg38 reference genome (and using the gencode V27 version of the genome annotation) using the STAR RNA-Seq alignment tool (Galaxy Version 2.5.2b-0) (Dobin et al., 2013). Low quality reads were removed by filtering (minimum MAPQ quality score = 30) (Galaxy Version 1.1.2) and count tables were generated with featureCounts (Galaxy Version 1.6.3) (Liao et al., 2014) and the gencode V27 version of the genome annotation. If indicated, RUVSeq was also utilised to remove unwanted variation from the RNA-seq data during the normalization of RNA-Seq read counts between samples (Risso et al., 2014). DESeq2 was utilised to identify differentially expressed genes (Love et al., 2014) and GOseq was utilised to perform Gene Ontology analysis (Young et al., 2010). PCA analysis was used to assess sample correlation. The analysis was carried out by Charles Girardot from Genome Biology Computational Support.

13.5 Single-cell RNA-sequencing via Drop-seq

13.5.1 Purification and cDNA synthesis

The library preparation for Drop-seq was carried out in a manner similar to what is outlined in Macosko et al., 2015. Starting with the generated emulsion (from Section 13.1.7), oil at the bottom of the tube was removed, before 30 ml of room temperature 6X SSC (Thermo Fisher Scientific Inc., USA) was added, and 1 ml of 1H,1H,2H,2H-Perfluorooctanol (PFO) (ABCR GmbH, Germany) added. The tube was then firmly inverted ten times to break the emulsion, before being spun at 1000 x *g* for 1 min at 4 °C, with reduced braking on the centrifuge.

The supernatant was carefully removed and 30 ml of 6X SSC was added to kick the beads into solution. Once a majority of the oil sank to the bottom, the supernatant was transferred to a new tube and spun at 1000 x *g* for 1 min at 4 °C. Once again, the supernatant was removed, leaving about 1 ml of liquid, which was then used to resuspend the beads for transfer to a DNA LoBind™ 1.5 ml tube (Eppendorf AG, Germany). The beads were spun at 1000 x *g* for 1 min at 4 °C and the supernatant was removed and discarded.

Three more washing steps were carried out in a similar manner - two with 6X SSC and one with 2X Maxima RT buffer (Thermo Fisher Scientific Inc., USA). 200 µl of RT mix was then added to the beads and they were incubated with rotation at room temperature for 30 min and then at 42 °C for 90 min. The beads were then washed once with TE-SDS, twice with TE-TW and once with 10 mm Tris pH 8.0, before being resuspended in 200 µl of exonuclease mix and incubated at 37 °C for 45 min with rotation. The beads were then washed once with TE-SDS, twice with TE-TW and once with nuclease-free water (Thermo Fisher Scientific Inc., USA), before being counted with a hemocytometer chamber.

2000 beads were mixed with PCR mix and PCR was carried out. PCR products were purified by the addition of 0.6x volume of SPRISelect Reagent (Beckman Coulter Inc., USA), before being subject to fragment size analysis using High Sensitivity DNA Chips and a 2100 Bioanalyzer (Agilent Technologies Inc., USA).

13.5.2 Tagmentation-based 3'-end library preparation

600 pg of purified cDNA was brought to a total volume of 5 µl with nuclease-free water, before 10 µl of Nextera TD buffer (Illumina Inc., USA) and 5 µl of Amplicon Tagment enzyme (Illumina Inc., USA) was added and the sample was mixed by pipetting and incubated at 55 °C for 5 min. 5 µl of Neutralisation Buffer (Illumina Inc., USA) was added and the solution was mixed by pipetting before incubation for 5 min at room temperature. The following reagents were then added to each PCR tube in this order: 15 µl of Nextera PCR mix (Illumina Inc., USA), 8 µl Nuclease-free water (Thermo Fisher Scientific Inc., USA), 1 µl 10 µM New-P5-SMART-PCR hybrid oligo and 1 µl 10 µM Nextera N70X oligo, before PCR was carried out with the Tagmentation PCR program.

PCR products were purified by the addition of 0.6X volume of SPRISelect Reagent (Beckman Coulter Inc., USA), before being subject to fragment size analysis using High Sensitivity DNA Chips and a 2100 Bioanalyzer (Agilent Technologies Inc., USA).

Samples were pooled at equimolar ratios and sequenced on a NextSeq 500 machine (Illumina Inc., USA), with paired-end sequencing carried out (Read 1: 20 bp, Read 2: 50 bp) and a Read 1 Index of 8 bp. A custom read 1 primer was used.

13.5.3 Data analysis

The library was first demultiplexed using the in-line i7 Illumina indices to assign reads to samples. Demultiplexing was carried out by Jonathan Landry from the EMBL Genomic Core Facility.

The dropSeqPipe pipeline was used for quality control, barcode detection, demultiplexing of droplets and for the generation of a UMI/read count table (Roelli, 2020).

The data was normalised by scran, prior to PCA analysis to assess sample correlation (Lun et al., 2016). scde was performed to identify differentially expressed genes (Kharchenko et al., 2014). The analysis was carried out by Charles Girardot from Genome Biology Computational Support.

13.6 Single-cell antibody sequencing via Drop-seq

13.6.1 Cell preparation and fluorescence activated cell sorting

Cells were washed twice in PBS, counted and resuspended in FACS sorting buffer at a concentration of 1 x 10^7 cells/ml. 4D5, H25B10 and OKT-9 cells were mixed at equal proportion and processed as a single sample, while anti-CD4 hybridoma cells were processed separately. DAPI (Thermo Fisher Scientific Inc., USA) was added to each sample to a final concentration of 100 nm.

Live cells were then sorted into 1.5 ml tubes with a BD FACSAria^{TM} Fusion Cell Sorter, utilising the 405 nm laser for the excitation of DAPI.

Sorted cells were spun down (5 min, 250 x *g*) at 4 °C, before the supernatant was removed and discarded and the cells were resuspended in PBS. This was repeated once for a total of two washing steps, before the cells were counted and resuspended in PBS with 0.01% Bovine Serum Albumin (BSA) (Sigma-Aldrich, USA) at a concentration of 4 x 10^6 cells/ml.

The 4D5, H25B10 and OKT-9 cell mixture and anti-CD4 hybridoma cells were then mixed at the desired ratios: 3:1 for the 1:1:1:1 sample and 1:16.67 for the 50:1:1:1 sample. For the 1:1:1+S sample, the 4D5, H25B10 and OKT-9 cell mixture was mixed with OKT-9 hybridoma supernatant and incubated for 20 min on ice, before being washed with PBS and resuspended in PBS with 0.01% Bovine Serum Albumin (BSA) (Sigma-Aldrich, USA) at a concentration of 4 x 10^6 cells/ml.

The samples were then processed via the Drop-seq encapsulation protocol as outlined in Section 13.1.7.

13.6.2 Purification and cDNA synthesis

The library preparation for Drop-seq was carried out in a manner similar to what is outlined in Macosko et al., 2015. Starting with the generated emulsion (from Section 13.1.7), oil at the bottom of the tube was removed, before 30 ml of room temperature 6X SSC (Thermo Fisher Scientific Inc., USA) was added, and 1 ml of 1H,1H,2H,2H-Perfluorooctanol (PFO) (ABCR GmbH, Germany) added. The tube was then firmly inverted ten times to break the emulsion, before being spun at 1000 x *g* for 1 min at 4 °C, with reduced braking on the centrifuge.

The supernatant was carefully removed and 30 ml of 6X SSC was added to kick the beads into solution. Once a majority of the oil sank to the bottom, the supernatant was transferred to a new tube and spun at 1000 x *g* for 1 min at 4 °C. Once again, the supernatant was removed, leaving about 1 ml of liquid, which was then used to resuspend the beads for transfer to a DNA LoBind^{TM} 1.5 ml tube (Eppendorf AG, Germany). The beads were spun at 1000 x *g* for 1 min at 4 °C and the supernatant was removed and discarded.

Three more washing steps were carried out in a similar manner - two with 6X SSC and one with 2X Maxima RT buffer (Thermo Fisher Scientific Inc., USA). 200 µl of RT mix was then added to the beads and they were incubated with rotation at room temperature for 30 min and then at 42 °C for 90 min. The beads were then washed once with TE-SDS, twice with TE-TW and once with 10 mm Tris pH 8.0, before being resuspended in 200 µl of exonuclease mix and incubated at 37 °C for 45 min with rotation. The beads were then washed once with TE-SDS, twice with TE-TW and once with nuclease-free water (Thermo Fisher Scientific Inc., USA), before being counted with a hemocytometer chamber.

10,000 beads were mixed with PCR mix for antibody sequencing and PCR was carried out. PCR products were purified by the addition of 0.6x volume of SPRISelect Reagent (Beckman Coulter Inc., USA), before the concentrations were measured by a Nanodrop spectrophotometer (Thermo Fisher Scientific Inc., USA) and Qubit fluorometer (Thermo Fisher Scientific Inc., USA) using the dsDNA High-Sensitivity Assay (as per the manufacturer's instructions) (Thermo Fisher Scientific Inc., USA). The samples were also subject to fragment size analysis using High Sensitivity DNA Chips and a 2100 Bioanalyzer (Agilent Technologies Inc., USA).

13.6.3 Single-molecule PCR amplification of antibody sequences

Purified PCR products obtained from the above PCR were encapsulated at a limiting dilution as described in Section 13.1.9, and subject to thermocycling as outlined in Section 13.15 ("Emulsion linkage PCR").

The thermocycled emulsion was combined in 2 ml DNA LoBind™ Tubes (Eppendorf AG, Germany). 0.8x volume of 1H,1H,2H,2H-Perfluorooctanol (PFO) (ABCR GmbH, Germany) was added to each tube and the mixture was pipetted up and down 10 - 15 times with a 1 ml pipette to break the emulsion. The tubes were then spun at 12,000 x *g* for 12 min at 4 °C. The supernatants were transferred and subject to PCR purification with MinElute columns (as per the manufacturer's instructions) (QIAGEN N.V., Netherlands), prior to two rounds of elution in 42 µl nuclease-free water (Thermo Fisher Scientific Inc., USA). PCR products were purified by the addition of 0.5x volume of AMPure XP beads (Beckman Coulter Inc., USA), and the concentrations were measured by a Nanodrop spectrophotometer (Thermo Fisher Scientific Inc., USA) and Qubit fluorometer (Thermo Fisher Scientific Inc., USA) using the dsDNA High-Sensitivity Assay (as per the manufacturer's instructions) (Thermo Fisher Scientific Inc., USA).

Nested PCR was carried out with 500 pg - 2 ng template. PCR products were then purified by the addition of 0.5x volume of AMPure XP beads (Beckman Coulter Inc., USA), and the concentrations were measured with a Qubit fluorometer (Thermo Fisher Scientific Inc., USA) using the dsDNA High-Sensitivity Assay (as per the manufacturer's instructions) (Thermo Fisher Scientific Inc., USA) before being

subject to fragment size analysis using High Sensitivity DNA Chips and a 2100 Bioanalyzer (Agilent Technologies Inc., USA).

Samples were pooled at equimolar ratios and sequenced on a MiSeq, with paired-end sequencing carried out (Read 1: 20 bp, Read 2: 236 bp, Read 3: 8 bp, Read 4: 236 bp). Custom read 1, read 2 and read 4 primers were used.

This protocol was established and optimised by Hongxing Hu and has been reproduced here with his permission.

13.6.4 Data analysis

The library was first demultiplexed using the in-line i7 Illumina indices to assign reads to samples. Demultiplexing was carried out by Jonathan Landry from the EMBL Genomic Core Facility.

Barcode detection and demultiplexing were carried out with the Je suite (Girardot et al., 2016). The paired-end reads (Read 2 and Read 4) corresponding to reads from the 3′ and 5′ ends of the antibody variable region were paired with PEAR (Paired-End reAd mergeR), with a maximum p-value of 0.01 and a minimum overlap of 10 bases (Zhang et al., 2014). Pear-merged antibody sequences were clustered by UMI to give a consensus sequence for each UMI, making use of VSEARCH (Rognes et al., 2016), using the centroid option with a minimum identity of 95%. The UMI-consensuses were then clustered by droplet barcodes, utilising VSEARCH (Rognes et al., 2016), using the centroid option with a minimum identity of 95%. IgBLAST was carried out for the droplet consensus sequences to identify the V, D and J regions (Ye et al., 2013). For the light chains, only the kappa chain sequences were included in the IgBLAST analysis. For the analysis of known antibody heavy and light chains, BLAST was used to align sequences against a custom database containing only the known antibody heavy and light chain sequences of 4D5, H25B10, anti-CD4 and OKT-9, which had been previously obtained with Sanger sequencing. Subsequent parsing and analyses were carried out in R. The analysis was carried out by Charles Girardot from Genome Biology Computational Support.

13.7 Single-cell sequencing via 10X Genomics

13.7.1 Cell hashing and 10X library preparation

Cells were prepared for Total-Seq A hashtag antibody staining as per the manufacturer's instructions (*Protocol - TotalSeq™-A Antibodies and Cell Hashing with 10x Single Cell 3′ Reagent Kit v3 3.1 Protocol*) (Biolegend, USA).

Adherent cells (MCF-7, A375-P, SK-BR-3) were harvested with 0.05% Trypsin-EDTA (Gibco, Thermo Fisher Scientific Inc., USA) before being washed twice with PBS and resuspended at a concentration of 2 x 10^7 cells/ml

in flow cytometry staining buffer. Suspension cells (K-562, Jurkat, U937) were washed twice with PBS and resuspended at a concentration of 2 x 10^7 cells/ml in flow cytometry staining buffer. 5 µl of Human TruStain FcX (Biolegend, USA) was added per 1 million cells and the samples were incubated at 4 °C for 10 min. At that time, 1 µg of hashtag antibody per sample was diluted in 50 µl flow cytometry staining buffer and spun at 14,000 x *g* at 4 °C for 10 min, before it was added to the samples and allowed to incubate at 4 °C for 30 min. 3.5 ml of flow cytometry staining buffer was then added to each sample and spun for 5 min at 250 x *g* at 4 °C, before the supernatant was removed and discarded. This washing step was repeated twice more. The cells were then resuspended in 400 µl PBS with 0.04% BSA (Sigma-Aldrich, USA), filtered through a 40 µm cell strainer (Corning Inc., USA) and counted with a TC10 Automated Cell Counter (Bio-Rad Laboratories Inc., USA). Cells were diluted to a final concentration of 1 x 10^6 cells/ml and equal volumes of the different samples were mixed.

Table 13.3: Antibodies utilised in cell hashing

Antibody name	**Hashtag oligo sequence**
Human TruStain FcX	-
TotalSeq™-A0251 anti-human Hashtag 1 Antibody	GTCAACTCTTTAGCG
TotalSeq™-A0252 anti-human Hashtag 2 Antibody	TGATGGCCTATTGGG
TotalSeq™-A0253 anti-human Hashtag 3 Antibody	TTCCGCCTCTCTTTG
TotalSeq™-A0254 anti-human Hashtag 4 Antibody	AGTAAGTTCAGCGTA
TotalSeq™-A0255 anti-human Hashtag 5 Antibody	AAGTATCGTTTCGCA
TotalSeq™-A0256 anti-human Hashtag 6 Antibody	GGTTGCCAGATGTCA
TotalSeq™-A0257 anti-human Hashtag 7 Antibody	TGTCTTTCCTGCCAG
TotalSeq™-A0258 anti-human Hashtag 8 Antibody	CTCCTCTGCAATTAC

All antibodies were obtained from Biolegend (Biolegend, USA).

Cells were subject to Chromium Single Cell 3' sample preparation as per the manufacturer's instructions (v3, CG000185 Rev B, or v3.1, CG000206 Rev D) (10X Genomics, USA). A Targeted Cell Recovery of 5000 cells was used for the experiments with K-562, Jurkat, MCF-7 and A375-P cells, while a Targeted Cell Recovery of 6000 cells was used for the experiment involving U937 and SK-BR-3 cells. At Step 2.2, cDNA amplification was carried out with the hashtag oligo (HTO) reaction mix, while the purification of cDNA and HTOs was done as outlined in the manufacturer's instructions (v3, CG000185 Rev B, or v3.1, CG000206 Rev D) (10X Genomics, USA).

The Cell Surface Protein Library Construction (Step 4) was not carried out, instead a Sample Index PCR Mix was prepared and amplified with the HTO Sample Index PCR protocol. The HTO library was then purified by the addition of 1.2X SPRISelect Reagent (Beckman Coulter Inc., USA). Both the cDNA and HTO libraries were then subject to fragment size analysis using High Sensitivity DNA Chips and a 2100 Bioanalyzer (Agilent Technologies Inc., USA). The libraries were pooled,

with the HTO library comprising 7% and the cDNA library comprising 93% for the experiments with K-562, Jurkat, MCF-7 and A375-P cells, and with the HTO library comprising 10% and the cDNA library comprising 90% for the experiments with U937 and SK-BR-3 cells. The libraries were sequenced on a NextSeq 500 machine (Illumina Inc., USA). Paired end libraries of cell barcodes (Read 1, 28 bp) and cDNA and HTOs (Read 2, 64 bp) were generated.

13.7.2 10X Genomics 3′ Gene Expression Analysis and cell hashing data analysis

Demultiplexing of the library was carried out with the Cell Ranger pipelines mkfastq and count (10X Genomics, USA). The Cell Ranger count pipeline was used to align reads to the GRCh38/hg19 reference genome, generate matrices of the number of reads per gene and to correlate cell barcodes with hashtag sequences. Demultiplexing and alignment of the sequencing results was done by Jonathan Landry from the EMBL Genomic Core Facility.

Seurat was utilised for quality control (QC) and normalisation of the data (Butler et al., 2018), where genes found in fewer than three cells and cells with fewer than 200 genes were removed. It was also used to identify highly variable features, cluster the cells and carry out dimensionality reduction and visualisation, by both t-distributed stochastic neighbourhood embedding (tSNE) (Maaten, 2014) and Uniform Manifold Approximation and Projection (UMAP) (McInnes et al., 2018). Default parameters were used, unless otherwise specified. Seurat analysis was carried out by Charles Girardot from Genome Biology Computational Support and by Xiaoli Ma from the Laboratory of Biomedical Microfluidics (LBMM), EPFL.

13.7.3 Subsampling analysis

Subsampling analysis was carried out on the data obtained from A375-P cells and unsynchronised MCF-7 cells, that were either treated with DMF (negative control), or treated with niclosamide. 763 A375-P cells treated with DMF (negative control) or 683 MCF-7 cells were included, together with a subset of randomly sampled treated cells. Prior to the analysis, where genes found in fewer than three cells and cells with fewer than 200 genes were removed. Clustering and dimensionality reduction were carried out via Seurat (Butler et al., 2018), and visualised with Uniform Manifold Approximation and Projection (UMAP) (McInnes et al., 2018). Default parameters were used, unless otherwise specified. This was repeated for different numbers of treated cells, namely 1, 2, 5, 10, 20, 30, 50, 100, 200, 300, 400 and 500 treated cells.

Subsampling analysis was carried out by Xiaoli Ma, Laboratory of Biomedical Microfluidics (LBMM), EPFL.

13.8 Flow cytometry

All flow cytometry analyses were conducted on a LSR Fortessa Analyser (Becton-Dickinson, USA), which is equipped with five lasers: 355 nm, 405 nm, 488 nm, 561 nm and 640 nm, for the simultaneous monitoring of various readouts.

13.8.1 Flow cytometry analysis of antibody expression of hybridoma cells

Cells were stained with anti-mouse IgG antibodies conjugated to Alexa Fluor 488 (Life Technologies, USA) prior to flow cytometry analysis on a LSR Fortessa Analyser (Becton-Dickinson, USA).

13.8.2 Flow cytometry analysis of CD14 expression on U937 cells

Increased CD14 expression was verified by staining U937 cells with mouse anti-human CD14 antibodies (Biolegend, USA) and anti-mouse IgG antibodies conjugated to Alexa Fluor 488 (Life Technologies, USA) prior to flow cytometry analysis on a LSR Fortessa Analyser (Becton-Dickinson, USA).

13.8.3 Flow cytometry analysis of sera binding to SK-BR-3 cells

Cells were stained with sera obtained from mice immunised with SK-BR-3 (Precision Antibody™, AG Pharmaceutical Inc., USA) and anti-mouse IgG antibodies conjugated to Alexa Fluor 568 (Life Technologies, USA) prior to flow cytometry analysis on a LSR Fortessa Analyser (Becton-Dickinson, USA).

13.8.4 Flow cytometry analysis of cell cycle synchronisation

Cells were resuspended in ice-cold PBS before ice-cold ethanol was added dropwise, to a final concentration of 80% ethanol. Samples were then incubated at -20°C for at least 3 hours. Cells were then centrifuged, before media removal and resuspension in PBS to enable cell rehydration. Cells were then resuspended in PBS containing 0.1% (v/v) Triton™ X-100 (Sigma-Aldrich, USA) and 1 µg/mL DAPI (Thermo Fisher Scientific Inc., USA), prior to flow cytometry analysis.

13.9 ELISA

Mouse IgG antigen detection via ELISA was carried out as per the manufacturer's instructions (Molecular Innovations Inc., USA or Abcam, United Kingdom). In each experiment, a standard curve was prepared for mouse IgG concentrations ranging from 0 to 500 ng/ml or 1000 ng/ml. Hybridoma samples were diluted between 10 and 200 times prior to the ELISA. Absorbance at 600 nm was measured in a Tecan Safire (Tecan Group Ltd., Switzerland) plate reader.

13.10 RNA extraction and reverse transcription

Cells were washed twice with PBS, before RNA extraction was carried out with the RNeasy Mini Kit (as per the manufacturer's instructions) (QIAGEN N.V., Netherlands). 250 ng of RNA was utilised for reverse transcription with SuperScript™ III Reverse Transcriptase (Thermo Fisher Scientific Inc., USA), according to manufacturer's instructions, utilising oligo $d(T)_{16}$ as the primers and an extension time of 45 min for the extension step at 50 °C. The resulting cDNA was diluted 20-fold prior to qPCR.

13.11 Real-time qPCR

Samples were mixed with the relevant primer pairs and SYBR Green PCR Master Mix (Applied Biosystems, Thermo Fisher Scientific Inc., USA), according to the manufacturer's instructions. The mixes were added to MicroAmp™ Fast Optical 96-Well Fast Clear Reaction Plates with Barcode (0.1 ml) (Applied Biosystems, Thermo Fisher Scientific Inc., USA) and run on a StepOne™ Real-Time PCR System (Applied Biosystems, Thermo Fisher Scientific Inc., USA).

Table 13.4: Primers utilised for qPCR

Gene name and primer pair	**Forward primer**	**Reverse primer**
INSIG1 pair 1	TACGCTGATCACGCAGTTTC	TGACGCCTCCTGAGAAAAAT
INSIG1 pair 2	CATTAACCACGCCAGTGCTA	CTGGAACGATCAAATGTCCA
ARHGEF2 pair 1	GTTTCAGGCATGACCATGTG	TGTCTTTACAGCGGTTGTGG
ARHGEF2 pair 2	TACCTGCGGCGAATTAAGAT	TGGGTCATCTCAGCAAACAG
MAT2A pair 1	GTTGTGCCTGCGAAATACCT	ATTTTGCGTCCAGTCAAACC
MAT2A pair 2	CTACGCCGTAATGGCACTTT	ACTCTGATGGGAAGCACAGC
TSC22D3 pair 1	GACCAGACCATGCTCTCCAT	CATGGCCTGTTCGATCTTGT
TSC22D3 pair 2	TGGTGGCCATAGACAACAAG	TGCTCCTTCAGGATCTCCAC
DDIT pair 1	CGAGTCCCTGGACAGCAG	TCACTGAGCAGCTCGAAGTC
DDIT pair 2	CGGAGGAAGACACGGCTTA	CATCAGGTTGGCACACAAGT
ACTB	TGAAGTGTGACGTGGACATC	GGAGGAGCAATGATCTTGAT
GAPDH	AGCAAGAGCACAAGAGGAAGAG	GAGCACAGGGTACTTTATTGATGG

All qPCR primers included are for the amplification of human genes

13.12 Sanger sequencing

Plasmid and PCR samples were prepared for Sanger sequencing as per the vendor's instructions, and sent to GATC for sequencing (Eurofins, Luxembourg).

13.13 Mouse immunisations

SK-BR-3 cell samples were harvested and aliquots of 2 million cells each were frozen at -80 °C in the presence of 10% fetal bovine serum (FBS) (Gibco, Thermo Fisher Scientific Inc., USA) and 10% DMSO.

13.13.1 In-house mouse immunisations

2 million frozen SK-BR-3 cells were thawed, washed with PBS and utilised for subcutaneous immunisation of female C57BL6 mice. The mice were boosted with 2 million SK-BR-3 cells at 4 and 8 weeks after the first immunisation. The animals were kept in the conventional containment conditions and fed *ad libitum*. Two weeks after the second booster, the mice were bled by retro-orbital route and sera were separated. The sera were tested with flow cytometry analysis.

13.13.2 Mouse immunisations and hybridoma generation by Precision Antibody

2 million frozen SK-BR-3 cells were thawed, washed with PBS and utilised for the immunisation of SJL/J mice, according to the company's protocols (Precision Antibody, AG Pharmaceutical Inc., USA), in the presence of adjuvant. A total of three mice were immunised. Sera were obtained via tailbleed and sent to EMBL for flow cytometry analysis.

Subsequently, mouse B-cells were obtained and fused with myeloma cells, with selection carried out via 1 - 2 weeks of post-fusion culture in HAT media, to remove unfused and dead myeloma and B-cells, all according to the company's protocols (Precision Antibody, AG Pharmaceutical Inc., USA). The cells were then frozen and shipped to EMBL on dry ice.

13.14 Buffers and solutions

13.14.1 Buffers

Table 13.5: Buffers

Buffer	Composition
Flow cytometry staining buffer	2% (v/v) FBS and 2 mm EDTA in PBS
FACS sorting buffer	2% (w/v) BSA and 2.5 mm EDTA in PBS
TE-SDS	10 mm Tris pH 8.0 with 0.5% SDS and 1 mm EDTA
TE-TW	10 mm Tris pH 8.0 with 0.01% Tween-20 and 1 mm EDTA

13.14.2 Reaction mixes

Table 13.6: Reaction mixes 1

Reaction mixes	Composition
HTO reaction mix	For 66 µl: 50 µl Amp Mix (10X Genomics, USA) 15 µl cDNA primers (10X Genomics 3' kit, 10X Genomics, USA) 1 µl HTO Additive Primer (0.2 µM stock)
Sample Index PCR Mix	For 100 µl: 5 µl Purified ADT/HTO fraction 2.5 µl SI PCR Primer (10 µM stock) (10X Genomics, USA) 2.5 µl TruSeq D701_S (10 µM stock) 50 µl 2X KAPA Hifi Master Mix (Sigma-Aldrich, USA) 40 µl Nuclease-free water (Thermo Fisher Scientific Inc., USA)
Lysis buffer	For 950 µl: 500 µl Nuclease-free water (Thermo Fisher Scientific Inc., USA) 300 µl 20% Ficoll PM-400 (GE Healthcare, USA) 10 µl 20% Sarkosyl (Sigma-Aldrich, USA) 40 µl 0.5 M EDTA (Thermo Fisher Scientific Inc., USA) 100 µl 2 M Tris pH 7.5 (Thermo Fisher Scientific Inc., USA)
RT mix	For 200 µl: 75 µl Nuclease-free water (Thermo Fisher Scientific Inc., USA) 40 µl Maxima 5X RT buffer (Thermo Fisher Scientific Inc., USA) 40 µl 20% Ficoll PM-400 (GE Healthcare, USA) 20 µl 10 mm dNTPs (VWR International, USA) 5 µl NXGen RNase Inhibitor (Lucigen, USA) 10 µl 50 µM Template Switch Oligo (TSO) (for Drop-seq) or TSO_P_M_LNA (for antibody sequencing) 10 µl Maxima h RTase (Thermo Fisher Scientific Inc., USA)
Exonuclease mix	For 200 µl: 20 µl Exo I buffer (New England Biolabs, USA) 170 µl Nuclease-free water (Thermo Fisher Scientific Inc., USA) 10 µl Exo I (New England Biolabs, USA)
Drop-seq PCR mix	For 50 µl, per 2000 beads: 24.6 µl Nuclease-free water (Thermo Fisher Scientific Inc., USA) 0.4 µl 100 µM TSO_PCR 25 µl 2X KAPA HiFi HotStart Readymix (Sigma-Aldrich, USA)

Table 13.7: Reaction mixes 2

Reaction mixes	Composition
Drop-seq PCR mix for antibody sequencing	For 50 µl, per 2000 beads: 24.2 µl Nuclease-free water (Thermo Fisher Scientific Inc., USA) 0.4 µl 50 µM SC_M_F_H 0.4 µl 50 µM SC_TSO_Rev_H 25 µl 2X KAPA HiFi HotStart Readymix (Sigma-Aldrich, USA)
Tagmentation mix	For 600 pg of cDNA in 5 µl water: 10 µl Nextera TD buffer (Illumina Inc., USA) 5 µl Amplicon Tagment enzyme (Illumina Inc., USA) *Incubate at 55 °C for 5 min* 5 µl Neutralisation Buffer (Illumina Inc., USA) *Incubate at room temperature for 5 min* 15 µl Nextera PCR mix (Illumina Inc., USA) 8 µl Nuclease-free water (Thermo Fisher Scientific Inc., USA) 1 µl 10 µM New-P5-SMART-PCR hybrid oligo 1 µl 10 µM Nextera N70X oligo
2X emulsion linkage PCR mix	For 1 ml of 2X PCR mix: 80 µl 10 µM 23-Abs-Common Forward 80 µl 10 µM 3-Abs-Common Reverse 4 µl 10 µM V5-Abs-Sense 4 µl H&L V-region Reverse Primer Mix, 10 µM overall 4 µl H&L V-region Sense Primer Mix, 10 µM overall 4 µl 10 µM Abs-PCR-BC 10 µl 50 mg/ml Ultrapure BSA (Thermo Fisher Scientific Inc., USA) 16 µl 25 mm dNTPs (VWR International, USA) 400 µl 5X Q5 reaction buffer (New England Biolabs, USA) 20 µl Q5 Hot Start High-Fidelity DNA Polymerase (New England Biolabs, USA) 368 µl Nuclease-free water (Thermo Fisher Scientific Inc., USA)
Nested PCR mix	For 20 µl: 0.4 µl V7-Mix-N50X Primer Mix, 10 µM overall 0.4 µl V7 Reverse Nested Primer Mix, 10 µM overall 0.8 µl Template DNA 8.4 µl Nuclease-free water (Thermo Fisher Scientific Inc., USA) 10 µl 2X KAPA HiFi HotStart ReadyMix (Sigma-Aldrich, USA)

13.15 PCR programmes

Table 13.8: PCR programmes 1

HTO Sample Index PCR		
Initial denaturation	98 °C	2 min
12 cycles	98 °C	20 s
	64 °C	30 s
	72 °C	20 s
Final extension	72 °C	5 min
Hold	4 °C	
PCR - Drop-seq		
Initial denaturation	95 °C	3 min
4 cycles	98 °C	20 s
	65 °C	45 s
	72 °C	3 min
9 cycles	98 °C	20 s
	67 °C	20 s
	72 °C	3 min
Final extension	72 °C	5 min
Hold	4 °C	
Tagmentation PCR		
Initial denaturation	95 °C	30 s
12 cycles	95 °C	10 s
	55 °C	30 s
	72 °C	30 s
Final extension	72 °C	5 min
Hold	4 °C	

Table 13.9: PCR programmes 2

Emulsion linkage PCR		
Initial denaturation	95 °C	30 sec
4 cycles	98 °C	5 s
	52 °C	30 s
	72 °C	30 s
4 cycles	98 °C	5 s
	55 °C	30 s
	72 °C	30 s
30 cycles	98 °C	5 s
	60 °C	30 s
	72 °C	45 s
Final extension	72 °C	2 min
Hold	4 °C	
Nested PCR		
Initial denaturation	95 °C	3 min
22 cycles	98 °C	20 s
	62 °C	20 s
	72 °C	30 s
Final extension	72 °C	5 min
Hold	4 °C	
RT-qPCR		
Initial denaturation	95 °C	10 min
40 cycles	95 °C	15 s
	60 °C	1 min
Melt curve stage	95 °C	15 s
	60 °C	1 min
Slow temperature increase	+ 0.3 °C	per min
	95 °C	15 s
End		

13.16 DNA sequences

Table 13.10: DNA sequences for Drop-seq and 10X Genomics

Sequence name	Sequence (5′ -> 3′)
HTO Additive primer	GTGACTGGAGTTCAGACGTGTGC*T*C
SI PCR primer	AATGATACGGCGACCACCGAGATCTACACTC TTTCCCTACACGACGC*T*C
TruSeq D701_S	CAAGCAGAAGACGGCATACGAGATCGAGTA ATGTGACTGGAGTTCAGACGTGT*G*C
TSO	AAGCAGTGGTATCAACGCAGAGTGAATrGrGrG
TSO_PCR	AAGCAGTGGTATCAACGCAGAGT
New-P5-SMART PCR Hybrid oligo	AATGATACGGCGACCACCGAGATCTACACG CCTGTCCGCGGAAGCAGTGGTATCAACGCAG AGT*A*C
Drop-seq custom Read 1 Primer	GCCTGTCCGCGGAAGCAGTGGTATCAACGCA GAGTAC

Table 13.11: DNA sequences for antibody sequencing PCR

Sequence name	Sequence (5′ -> 3′)
TSO_P_M_LNA	AAGCAGTGGTATCAACGCAGAGTGAATTCGC GGCCGCTCGCGAGAATrGrG+G
SC_M_F_H	GTGAATTCGCGGCCGCTCGCGAG
SC_TSO_Rev_H	AAGCAGTGGTATCAACGCAGAGTAC
23-Abs-Common Forward	CGCAGTAGCGGTAAACGGC
3-Abs-Common Reverse	GCGGATAACAATTTCACACAGGS
V5-Abs-Sense	GCGCCGCGATGGGAATAGTGAATTCGCGGCC GCTCGCGAG
Abs-PCR-BC	TATTCCCATCGCGGCGCAAGCAGTGGTATCA ACGCAGAGTAC

Table 13.12: H&L V-region Reverse Primer Mix

Sequence name	Sequence (5′ -> 3′)	Proportion to add
V7-Abs-Ighg12abc	CGCAGTAGCGGTAAACGGCCTGGACAGGG ATCCAGAGTTCC	18
V7-Abs-Ighg3	CGCAGTAGCGGTAAACGGCCTGGACAGGG CTCCATAGTTCC	2
V7-Abs-Igkc	CGCAGTAGCGGTAAACGGCCGACTGAGGC ACCTCCAGATGTTAACTGCTC	18
V7-Abs-Iglc23	CGCAGTAGCGGTAAACGGCCCCTGGGTGAT AGGTGTACCATTTGC	1
V7-Abs-Iglc14	CGCAGTAGCGGTAAACGGCCCCTGAGTGA MAGGGGTACCATCTRC	1

Table 13.13: H&L V-region Sense Primer Mix

Sequence name	Sequence (5′ -> 3′)	Proportion to add
4-Abs-P1-BC-Ighg1	GCGGATAACAATTTCACACAGGACCTGCTC TGTGTTACATGAGGGCCTGCAC	36
4-Abs–P1-BC-Ighg2a	GCGGATAACAATTTCACACAGGGTTCAGTG GTCCACGAGGGTCTGCACAATCACC	1
4-Abs–P1-BC-Ighg2b	GCGGATAACAATTTCACACAGGCCTTCTCAT GCAACGTGAGACACGAGGGTCTG	1
4-Abs–P1-BC-Ighg2c	GCGGATAACAATTTCACACAGGGCCTGCTC AGTGGTCCACGAGGGTCTGC	1
4-Abs–P1-BC-Ighg3	GCGGATAACAATTTCACACAGGACCTGCTC CGTGGTGCATGAGGCTCTCC	1
4-Abs-P1-BC-Igkc	GCGGATAACAATTTCACACAGGGAGGCCAC TCACAAGACATCAACTTCACCCATTGTC	36
4-Abs-P1-BC-Iglc1/2/3/4	GCGGATAACAATTTCACACAGGGMCTGGA AGGYARATGGTACMCCTDTCACYCAGG	4

Table 13.14: V7-Mix-N50X Primer Mix

Sequence name	Sequence (5′ -> 3′)	Proportion to add
V72-Nest-Igh12abc-50X	AATGATACGGCGACCACCGAGATCTACACx xxxxxxxCAGGGGCCAGTGGATAGACHGATG GGG	18
V72-Nest-Igh3-50X	AATGATACGGCGACCACCGAGATCTACACx xxxxxxxCAGCCAGGGACCAAGGGATAGACA GATG	2
V7-Nest-K-50X	AATGATACGGCGACCACCGAGATCTACACx xxxxxxxGCTCACTGGATGGTGGGAAGATGGA TACAGTTGG	18
V72-Nest-L14-50X	AATGATACGGCGACCACCGAGATCTACACx xxxxxxxGAGCTCTTCAGAGGAAGGTGGRAAC AG	1
V72-Nest-L23-50X	AATGATACGGCGACCACCGAGATCTACACx xxxxxxxGAGCTCCTCAGRGGAAGGTGGAAAC A	1

xxxxxxxx refer to the 8bp adaptor sequence (refer to Table 13.15)

Table 13.15: Adaptor sequences

Index name	Adaptor sequence (5′ -> 3′)
501	TAGATCGC
502	CTCTCTAT
503	TATCCTCT
504	AGAGTAGA
505	GTAAGGAG
506	ACTGCATA
507	AAGGAGTA

Table 13.16: V7 Reverse Nested Primer Mix

Sequence name	Sequence (5′ -> 3′)	Proportion to add
V7-Abs-BC-Ighg1	CAAGCAGAAGACGGCATACGAGATGAGGG CCTGCACAACCACCATACTGAGAAGAGC	36
V7-Abs–BC-Ighg2a	CAAGCAGAAGACGGCATACGAGATGAGGG TCTGCACAATCACCACACGACTAAGAGCTT CTCC	1
V7-Abs–BC-Ighg2b	CAAGCAGAAGACGGCATACGAGATGAGGG TCTGAAAAATTACTACCTGAAGAAGACCAT CTCCCGG	1
V7-Abs–BC-Ighg2c	CAAGCAGAAGACGGCATACGAGATGAGGG TCTGCACAATCACCTTACGACTAAGACCAT CTCC	1
V7-Abs–BC-Ighg3	CAAGCAGAAGACGGCATACGAGATGAGGC TCTCCATAACCACCACACACAGAAGAACCT GTCTC	1
V7-Abs–BC-Igkc	CAAGCAGAAGACGGCATACGAGATCTTCAC CCATTGTCAAGAGCTTCAACAGGAATGAGT G	36
V7-Abs–BC-Iglc1234	CAAGCAGAAGACGGCATACGAGATCADGT YACWCATGAAGGKVACACTGTGGAGAAGA GT	4

Table 13.17: Sequencing primers for antibody sequencing

Sequence name	Sequence (5′ -> 3′)
Read 1	CCATCGCGGCGCAAGCAGTGGTATCAACGCA GAGTAC
Read 2	Read 2 Sequencing Primer Mix (refer to Table 13.18)
Read 4	GAATTCGCGGCCGCTCGCGAGAATGGG

Table 13.18: Read 2 Sequencing Primer Mix

Sequence name	Sequence (5′ -> 3′)	Proportion to add
SC-Ighg1	GGGGCCAGTGGATAGACAGATGGGGGTGTC GTTTTGG	36
SC-Ighg2a	GGCCAGTGGATAGACCGATGGGGCTGTTGT TTTGGC	1
SC-Ighg2b-v7	GGGGCCAGTGGATAGACTGATGGGGGTGTT GTTTTGG	1
SC-Ighg2c	GGGGCCAGTGGATAGACCGATGGGGCTGTT GTTTTGG	1
SC-Ighg3	CAGCCAGGGACCAAGGGATAGACAGATGG GGCTGTTGTTGTAG	1
SC-Igkc-nu	TGGTGGGAAGATGGATACAGTTGGTGCAGC ATCAGCCC	36*
SC-Igkc-Iso	GCTCACTGGATGGTGGGAAGATGGATACAG TTGGTGCAGCATC	36*
SC-Iglc1	GAAACAGGGTGACTGATGGCGAAGACTTGG GCTGGC	1
SC-Iglc2	GGAAACACGGTGAGAGTGGGAGTGGACTT GGGCTGAC	1
SC-Iglc3	TGGAAACATGGTGAGTGTGGGAGTGGACTT GGGCTGAC	1
SC-Iglc4	AGGTGGGAACAGATTAACTGAGGGTGTAGC CTTGGGTTGGC	1

*Add either SC-Igkc-nu or SC-Igkc-Iso for the sequencing of the kappa chain

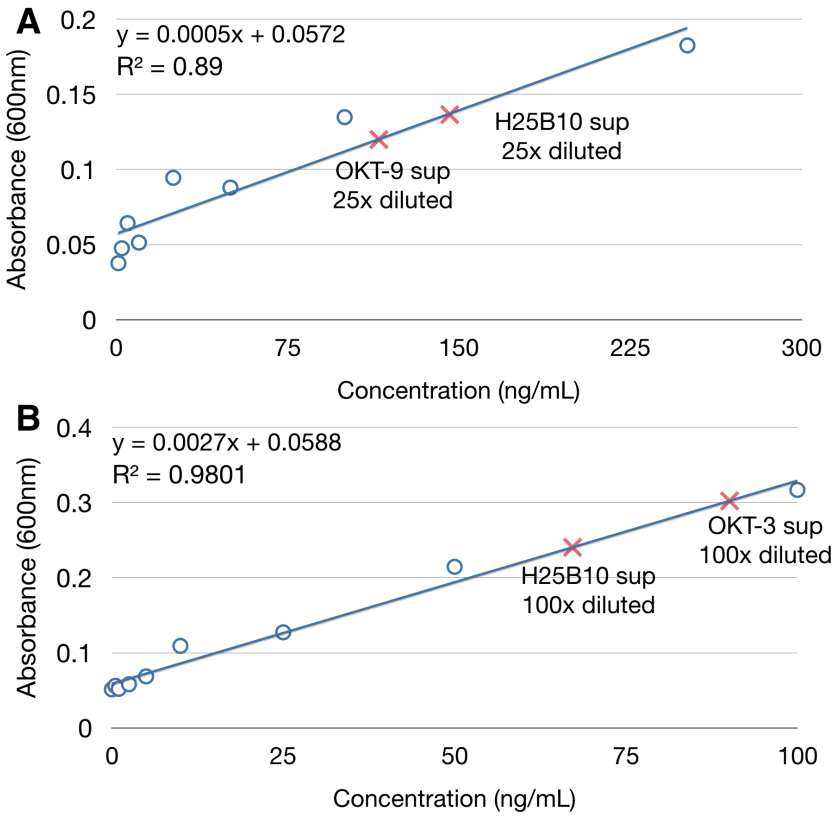
Table 13.19: DNA sequence modifications

Symbol	Modification
*	Phosphorothioated base
r	RNA base
+	Locked nucleic acid (LNA)

A Determining antibody concentrations of hybridoma supernatants

Enzyme-linked immunosorbent assays (ELISAs) were carried out to accurately assess the antibody concentrations present in hybridoma supernatants used for stimulation experiments.

A.1 OKT-9, OKT-3 and H25B10 hybridoma supernatants



(see next page)

Figure A.1: Assessing mouse IgG concentrations of OKT-9, OKT-3 and H25B10 hybridoma supernatants. Standard curves were independently prepared for each experiment (blue lines) based on the absorbance values at 600 nm obtained for mouse IgG standards (hollow blue circles). Absorbance values at 600 nm are shown on the y-axis while mouse IgG concentrations in ng/ml are shown on the x-axis. **A.** Mouse IgG concentrations of 25x diluted OKT-9 and H25B10 supernatants were assessed (red crosses) prior to incubation with K562 cells. **B.** Mouse IgG concentrations of 100x diluted OKT-3 and H25B10 supernatants were assessed (red crosses) prior to incubation with Jurkat cells. Data was generated with the mouse Immunoglobulin G ELISA Kit from Molecular Innovations.

The mouse IgG concentrations of the supernatants used for incubation with K-562 cells were found to be 3.14 µg/ml and 3.97 µg/ml for OKT-9 and H25B10 respectively (Fig. A.1A), while those used for incubation with Jurkat cells had concentrations of 9.01 µg/ml and 6.72 µg/ml for OKT-3 and H25B10 respectively (Fig. A.1B). In each experiment, the supernatants were diluted to the lower concentration to ensure that an equal amount of antibody was added to each sample.

A.2 60bca and H25B10 hybridoma supernatants

The mouse IgG concentrations of the 60bca and H25B10 hybridoma supernatants were 5.36 µg/ml and 7.01 µg/ml respectively (Fig. A.2). In each experiment, the supernatants were diluted to 5.36 µg/ml to ensure that an equal amount of antibody was added to each sample.

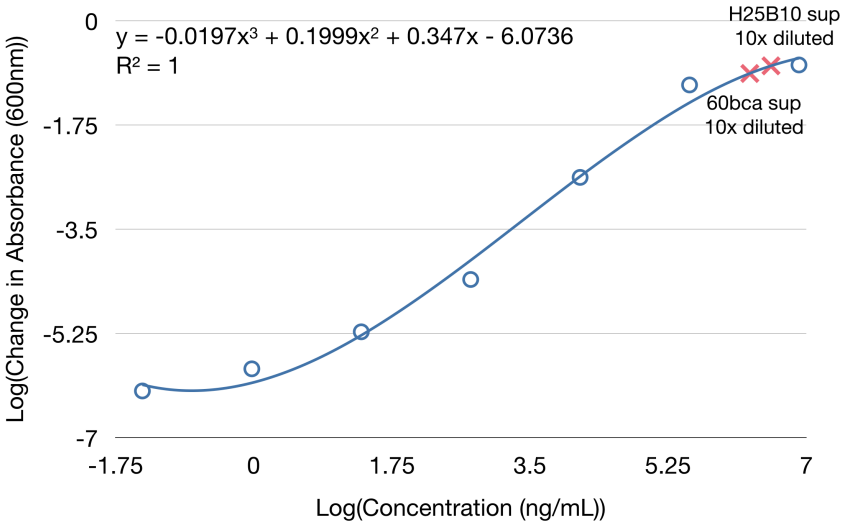


Figure A.2: Assessing mouse IgG concentrations of 60bca and H25B10 hybridoma supernatants. A standard curve was prepared (blue line) based on the changes in absorbance values at 600 nm obtained for mouse IgG standards over 10 minutes (hollow blue circles). The natural logarithms of the change in absorbance values at 600 nm are shown on the y-axis while the natural logarithms of mouse IgG concentrations in ng/ml are shown on the x-axis. Mouse IgG concentrations of 10x diluted 60bca and H25B10 supernatants were assessed (pink crosses) prior to incubation with differentiated U937 cells. Data was generated with the IgG mouse ELISA Kit from abcam.

A.3 8G5F11, IE9F9 and H25B10 hybridoma supernatants

The mouse IgG concentrations of the 8G5F11, IE9F9 and H25B10 hybridoma supernatants were found to be 7.74 µg/ml, 12.09 µg/ml and 24.72 µg/ml respectively (Fig. A.3). The hybridoma supernatants were all diluted to 7.5 µg/ml prior to use, to ensure that an equal amount of antibody was added to each sample.

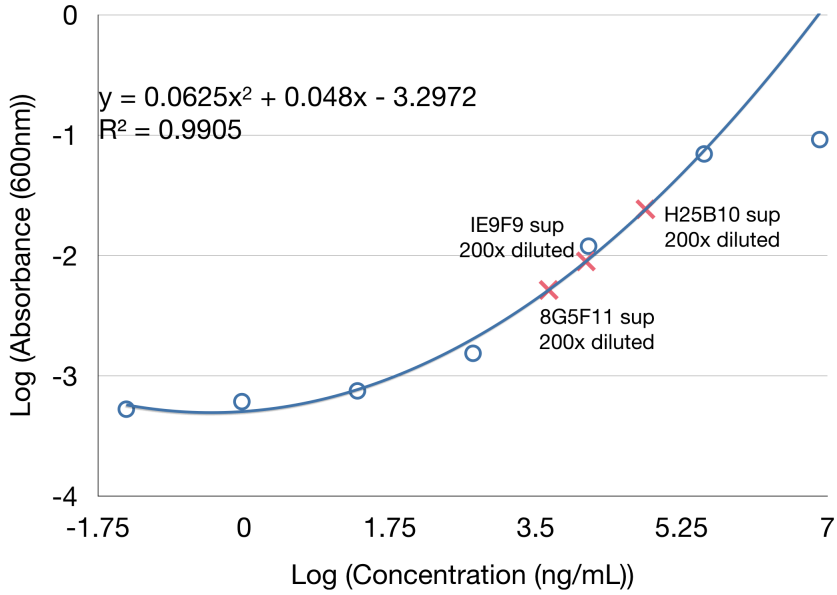


Figure A.3: Assessing mouse IgG concentrations of 8G5F11, IE9F9 and H25B10 hybridoma supernatants. A standard curve was prepared (blue line) based on the absorbance values at 600 nm obtained for mouse IgG standards (hollow blue circles). The natural logarithms of the absorbance values at 600 nm are shown on the y-axis while the natural logarithms of mouse IgG concentrations in ng/ml are shown on the x-axis. Mouse IgG concentrations of 200x diluted 8G5F11, IE9F9 and H25B10 hybridoma supernatants were assessed (pink crosses) prior to incubation with lentiviruses. Data was generated with the IgG mouse ELISA Kit from abcam.

A.4 Precision Antibody and H25B10 hybridoma supernatants

The mouse IgG concentrations of the Precision Antibody and H25B10 hybridoma supernatants were found to be 15.7 µg/ml and 15.8 µg/ml respectively (Fig. A.4). The H25B10 supernatant was diluted to 15.7 µg/ml prior to use, to ensure that an equal amount of antibody was added to each sample.

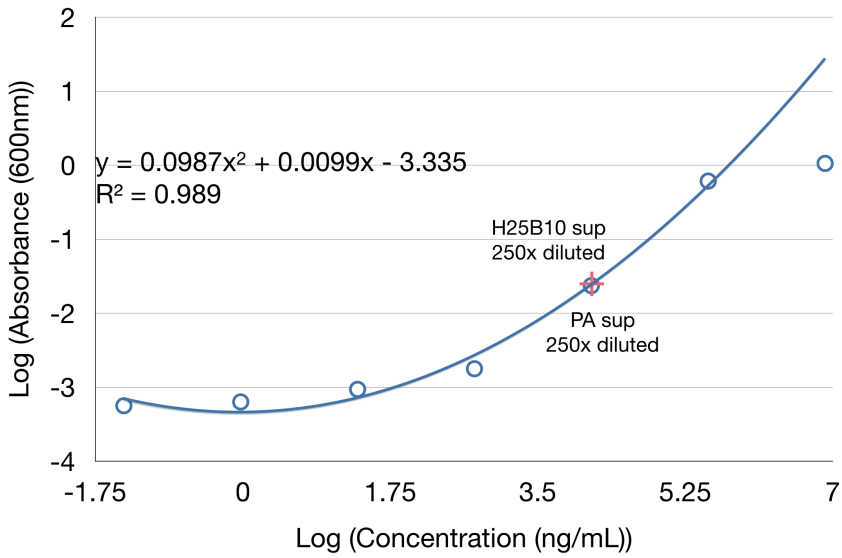
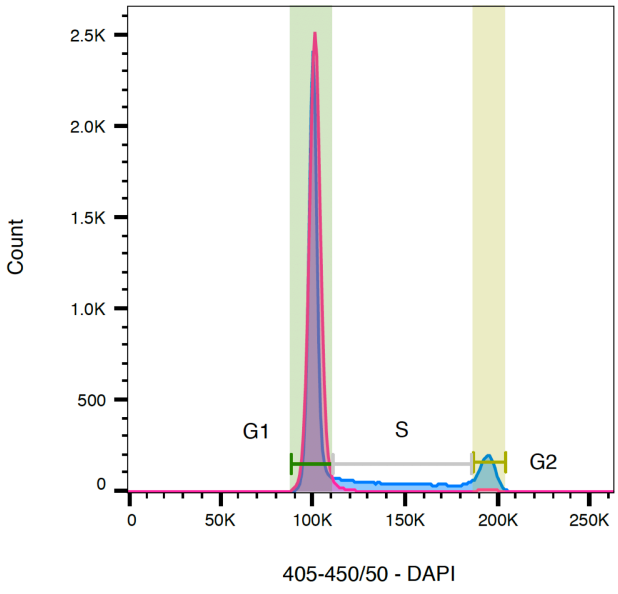


Figure A.4: Assessing mouse IgG concentrations of Precision Antibody and H25B10 hybridoma supernatants. A standard curve was prepared (blue line) based on the absorbance values at 600 nm obtained for mouse IgG standards (hollow blue circles). The natural logarithms of the absorbance values at 600 nm are shown on the y-axis while the natural logarithms of mouse IgG concentrations in ng/ml are shown on the x-axis. Mouse IgG concentrations of 250x diluted Precision Antibody and H25B10 hybridoma supernatants were assessed (pink crosses). Data was generated with the IgG mouse ELISA Kit from abcam.

B | Cell cycle synchronisation with palbociclib

Palbociclib is an inhibitor of the phosphorylation of CDK4/6 (Finn et al., 2009). This results in hypophosphorylation of pRb, preventing S-phase entry and the DNA synthesis required for cell division, therefore arresting the cell cycle at the G1 phase.



	Sample Name	G1 :: Freq. of Parent	S :: Freq. of Parent	G2 :: Freq. of Parent
	MCF-7 + 2.5 µM palbociclib	95.9	2.93	1.15
	MCF-7 + DMSO	69.0	19.8	11.0

Figure B.1: Cell cycle synchronisation of MCF-7 cells with 2.5 µM palbociclib. Cells were dehydrated by incubation in 80% ethanol at -20 °C for at least 3 hours. They were then rehydrated and resuspended in PBS containing 0.1% (v/v) Triton™ X-100 and 1 µg/mL DAPI, prior to flow cytometry analysis. The cells present in the G1 (green), S (grey) and G2 (mustard) phases can be distinguished by their DAPI signal, which indicates the relative amount of DNA present in the cells. DAPI was detected at 405-450/50, X-axis.

We have established cell cycle synchronisation in MCF-7 cells. In unsynchronised MCF-7 cells, 31% of cells were in the S and G2 phases, with 69% of cells in the G1 phase (Fig.B.1A, blue), while in the presence of 2.5 µM palbociclib, 4%

of cells were in the S and G2 phases and 96% of cells were in the G1 phase (Fig.B.1A, pink).

C | Top differentially expressed genes in sequencing experiments

Gene Symbol	Log2FC	Wald-Stats	p-value	Adj p-value
SOD2	2.26100099	96.8340309	0	0
FTH1	1.83301037	77.2526022	0	0
CCL2	2.62528497	57.1850985	0	0
ACSL1	1.20570079	54.1249997	0	0
CXCL8	2.03024681	52.4335506	0	0
SLC7A11	1.51443829	43.3581112	0	0
OLIG1	2.31657488	43.339373	0	0
NFKB2	1.49419295	43.0426866	0	0
TNFAIP6	2.5209804	41.8588466	0	0
PIK3AP1	1.10339892	40.4243299	0	0
CD44	1.17950928	38.2043166	0	0
TNFAIP3	1.64623616	37.5887004	0	0
HCK	1.10414668	37.4613289	3.93E-307	4.09E-304
SERPINB2	1.17650618	37.4596762	4.18E-307	4.09E-304
FAM129A	1.00603981	36.7971634	2.05E-296	1.87E-293
GAS7	0.86924842	36.3218309	7.32E-289	6.26E-286
SAMSN1	1.31478117	33.7317769	1.98E-249	1.59E-246
CLIC4	0.8637671	32.9207563	1.11E-237	8.44E-235
ATP2B1	1.06900173	32.8380938	1.68E-236	1.21E-233
C3	2.26722232	32.7286807	6.11E-235	4.18E-232

Table C.1: Top 20 differentially expressed genes in bulk RNA-seq of CD14-expressing U937 cells upon LPS activation. The genes shown are differentially expressed in bulk RNA-seq data of CD14-expressing U937 cells upon LPS activation in the presence of H25B10 hybridoma supernatant, compared to CD14-expressing U937 cells upon LPS activation in the presence of 60bca hybridoma supernatant. Data reproduced with the permission of Charles Girardot, EMBL Heidelberg.

Gene Symbol	Log2FC	Wald-Stats	p-value	Adj p-value
TurboGFP	5.62258704	190.7584374	0	0
Soga3	4.93115507	46.3402174	0	0
Il11	4.39984653	47.5338354	0	0
Clcf1	3.46346797	43.0567027	0	0
Thbd	3.15006871	98.9401012	0	0
Tfap2a	2.46951645	59.3494976	0	0
Gclc	1.98247736	80.6964055	0	0
Adamts1	1.88556156	41.8020556	0	0
Phlda1	1.83994729	41.6610294	0	0
PEX5	1.66828033	56.397473	0	0
Gata2	1.60672452	41.7900229	0	0
Sgk1	1.57710398	38.5448287	0	0
Dusp4	1.56705335	49.7296723	0	0
Sh3bp5	1.52927531	41.9652613	0	0
Scarb1	1.28906643	46.3408639	0	0
Tnxb	1.28888809	39.0513561	0	0
Errfi1	1.23050478	38.4015003	0	0
Net1	1.22647673	42.6468222	0	0
Rrm2	1.20722909	53.0336263	0	0
Scd2	1.19243674	37.5545413	0	0
Tsc22d1	1.17602778	46.0975193	0	0
Lif	1.09653009	43.5121945	0	0
Igfbp4	0.66641723	57.4240046	0	0
Strip2	-1.8330434	-44.311021	0	0

Table C.2: Top 24 differentially expressed genes in bulk RNA-seq of CHO-GLP1R-GFP cells upon addition of 10 µM Exendin-4. The genes shown are differentially expressed in bulk RNA-seq data of CHO-GLP1R-GFP cells upon incubation with 10 µM Exendin-4 for 6 hours, compared to untreated CHO-GLP1R-GFP cells. Data reproduced with the permission of Charles Girardot, EMBL Heidelberg.

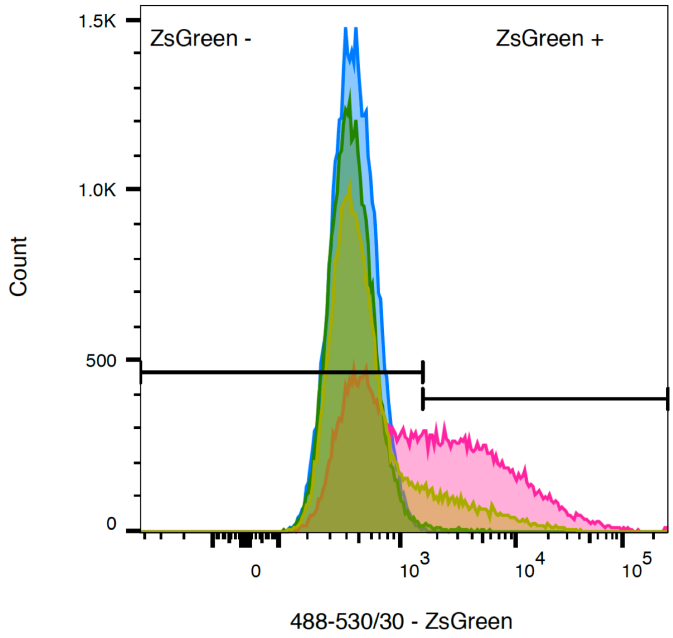
Gene symbol	Log2FC	Z-score adj	p-value	Adj p-value
Fam188b	8.24804448	6.25304523	8.02E-13	4.03E-10
Smox	6.98423122	6.25304523	9.61E-13	4.03E-10
Serpinf1	6.61839053	6.25304523	8.05E-13	4.03E-10
TurboGFP	4.75592888	6.25304523	8.02E-13	4.03E-10
Thbd	2.26156058	6.25304523	8.02E-13	4.03E-10
Cited1	2.06201112	6.25304523	8.02E-13	4.03E-10
Scarb1	1.72942868	6.25304523	8.02E-13	4.03E-10
Tfap2a	1.56313746	6.25304523	8.19E-13	4.03E-10
Pbx1	1.46336273	6.25304523	8.02E-13	4.03E-10
Rab31	1.43010449	6.25304523	8.06E-13	4.03E-10
Timp1	1.23055502	6.25304523	8.59E-13	4.03E-10
Rplp1	1.23055502	6.25304523	8.02E-13	4.03E-10
Itgb5	1.0642638	6.25304523	8.13E-13	4.03E-10
Lrpap1	0.93123083	6.25304523	9.40E-13	4.03E-10
Lgals1	0.86471434	6.25304523	8.02E-13	4.03E-10
Gpx4	0.79819785	6.25304523	8.02E-13	4.03E-10
Arpc1b	0.76493961	6.25304523	8.02E-13	4.03E-10
Id3	0.73168137	6.25304523	8.73E-13	4.03E-10
Ankrd1	-2.1950441	-6.2530452	8.02E-13	4.03E-10

Table C.3: Top 19 differentially expressed genes in Drop-seq of CHO-GLP1R-GFP cells upon addition of 10 µM Exendin-4. The genes shown are differentially expressed in Drop-seq data of CHO-GLP1R-GFP cells upon incubation with 10 µM Exendin-4 for 24 hours, compared to untreated CHO-GLP1R-GFP cells. Data reproduced with the permission of Charles Girardot, EMBL Heidelberg.

D Inhibition of lentiviral transduction

We have tested antibodies secreted by two hybridoma cell lines, namely 8G5F11 and IE9F9, for their ability to inhibit lentiviral transduction (Lefrancois and Lyles, 1982; Munis et al., 2018). This has been carried out for our cell line of interest, K-562, while utilising a lentivirus at a multiplicity of infection (MOI) of 5. This lentivirus was generated using the transfer plasmid pLVX-IRES-ZsGreen, such that the lentiviral particles contained a gene encoding a protein expressing green fluorescence (ZsGreen). Lentiviral transduction in the presence of 7.5 µg/mL H25B10 hybridoma supernatant (negative control) resulted in 47.3% of cells being ZsGreen+ (Fig. D.1, pink). The ZsGreen+ population of K-562 cells was reduced to 1.34% and 14.3% respectively, in the presence of 7.5 µg/mL 8G5F11 and IE9F9 hybridoma supernatants (Fig. D.1, green, mustard), illustrating that the anti-VSV-G antibodies present inhibited the lentiviral transduction of K-562 cells by lentiviral particles containing the gene for ZsGreen. In addition, the 8G5F11 antibody was shown to be more effective than the IE9F9 antibody at inhibiting lentiviral transduction.

Antibody concentrations of hybridoma supernatants were determined by ELISA (Appendix A).



	Sample Name	ZsGreen - :: Freq. of Parent	ZsGreen + :: Freq. of Parent
	Virus + IE9F9 supernatant	85.7	14.3
	Virus + 8G5F11 supernatant	98.7	1.34
	Virus + H25B10 supernatant	52.7	47.3
	No virus	99.8	0.16

Figure D.1: Inhibition of lentiviral transduction by anti-VSV-G hybridoma supernatants. Lentivirus containing the gene for ZsGreen were mixed with 7.5 µg/mL hybridoma supernatant (H25B10: pink, 8G5F11: green, IE9F9: mustard) and incubated at 37 °C for 1 hour. The lentivirus-antibody mixture was incubated with K-562 cells overnight. K-562 cells were analysed by flow cytometry 96 hours after viral transduction. A sample where no lentivirus and no hybridoma supernatant was added was included as a negative control (blue). ZsGreen expression was detected at 488-530/30, X-axis.

E | Distribution of UMIs in droplets during antibody sequencing

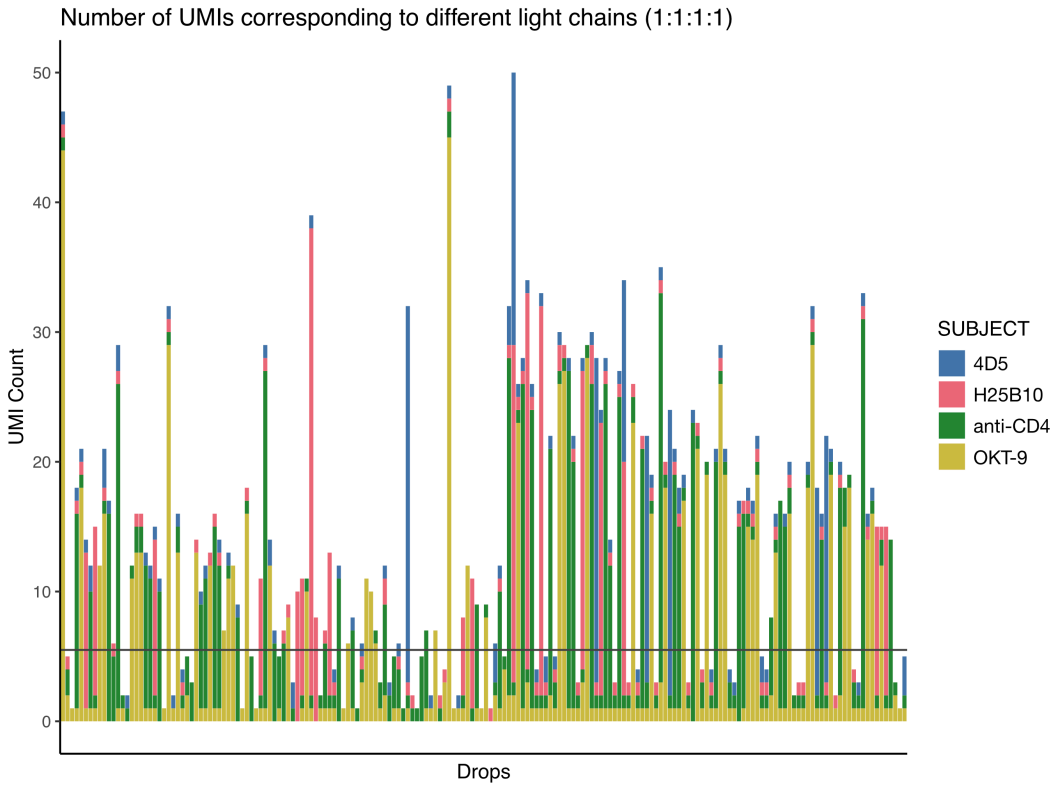


Figure E.1: Distribution of UMIs associated with droplets for kappa chain sequencing of the 1:1:1:1 hybridoma mixture. Four hybridoma cell lines (4D5, H25B10, anti-CD4 and OKT-9) were mixed in equal proportion. Independent droplets are shown on the x-axis ("Drops") while the UMI count for each droplet is depicted on the y-axis ("UMI Count"). Each bar corresponds to the UMIs associated within a single droplet. The identity of the kappa chain associated with each UMI is indicated by the colour(s) of the bar ("SUBJECT", with 4D5: blue, H25B10: pink, anti-CD4: green, OKT-9: mustard). The grey line at UMI Count = 5.5 indicates the cut-off utilised. Droplets with bars above this cut-off were further processed, while droplets with bars below this, which have five or fewer UMIs associated to identifiable kappa chains, were rejected.

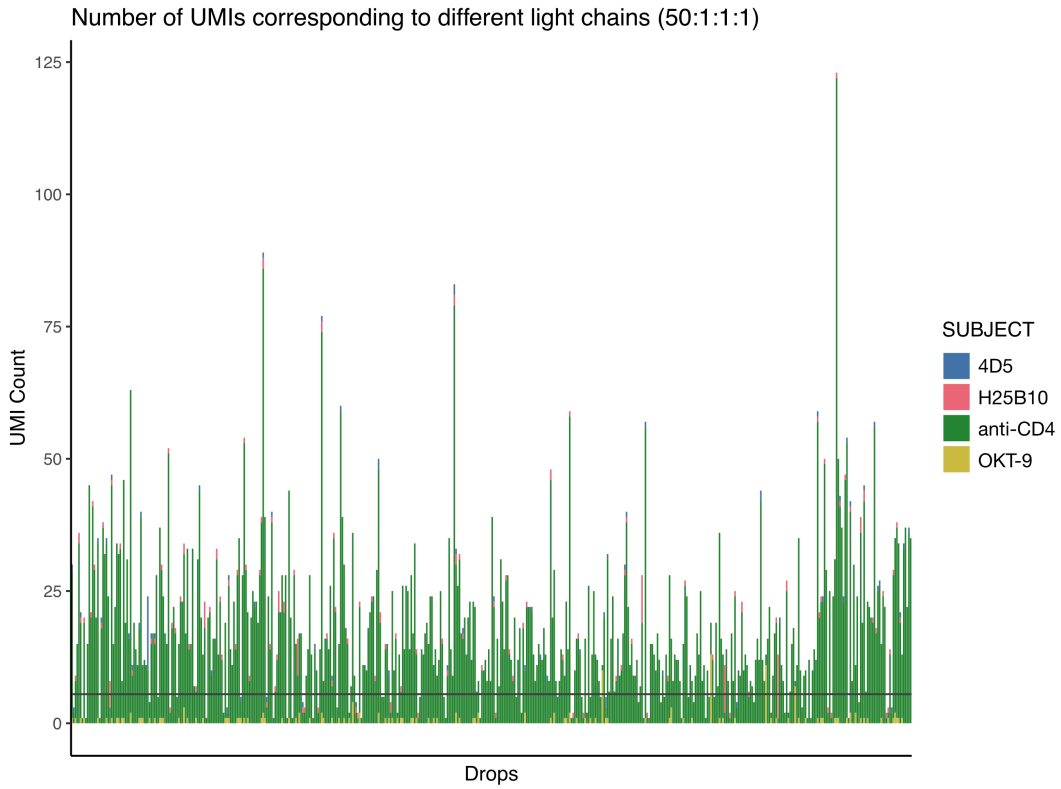


Figure E.2: Distribution of UMIs associated with droplets for kappa chain sequencing of the 50:1:1:1 hybridoma mixture. Three hybridoma cell lines (4D5, H25B10 and OKT-9) were mixed in equal proportion with anti-CD4 hybridoma cells added at 50-fold the concentration of the other cell lines. Independent droplets are shown on the x-axis ("Drops") while the UMI count for each droplet is depicted on the y-axis ("UMI Count"). Each bar corresponds to the UMIs associated within a single droplet. The identity of the kappa chain associated with each UMI is indicated by the colour(s) of the bar ("SUBJECT", with 4D5: blue, H25B10: pink, anti-CD4: green, OKT-9: mustard). The grey line at UMI Count = 5.5 indicates the cut-off utilised. Droplets with bars above this cut-off were further processed, while droplets with bars below this, which have five or fewer UMIs associated to identifiable kappa chains, were rejected.

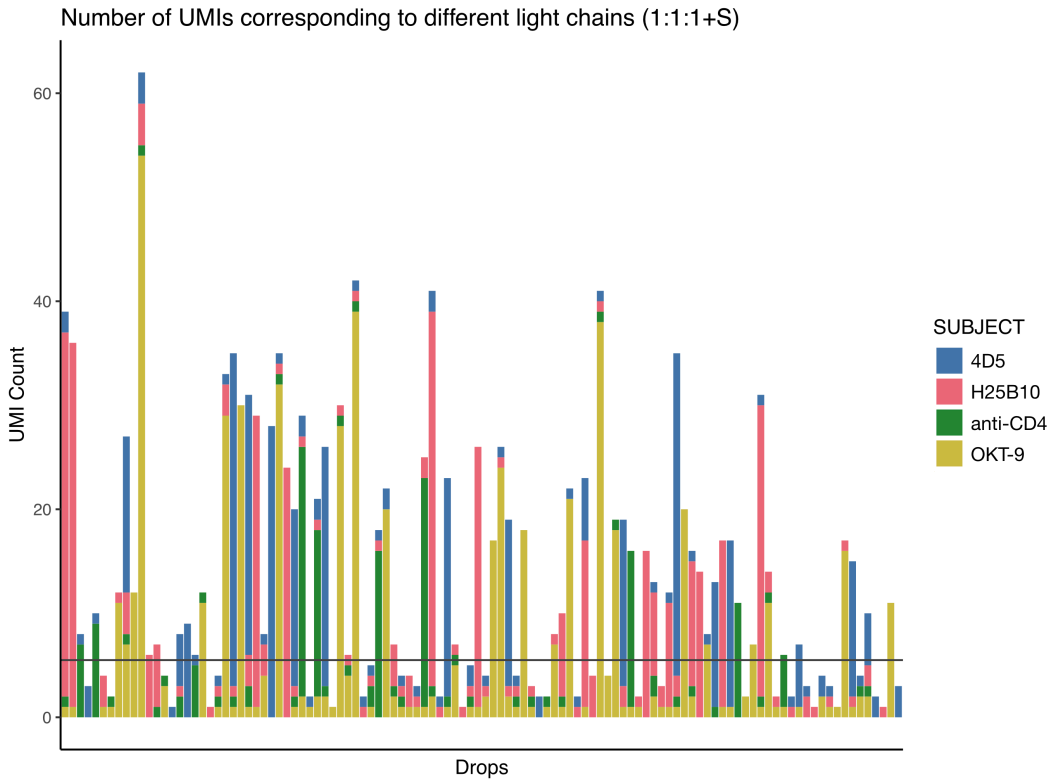


Figure E.3: Distribution of UMIs associated with droplets for kappa chain sequencing of the 1:1:1+S hybridoma mixture. Three hybridoma cell lines (4D5, H25B10 and OKT-9) were mixed in equal proportion, and incubated with anti-CD4 supernatant for 20 minutes, prior to processing. Independent droplets are shown on the x-axis ("Drops") while the UMI count for each droplet is depicted on the y-axis ("UMI Count"). Each bar corresponds to the UMIs associated within a single droplet. The identity of the kappa chain associated with each UMI is indicated by the colour(s) of the bar ("SUBJECT", with 4D5: blue, H25B10: pink, anti-CD4: green, OKT-9: mustard). The grey line at UMI Count = 5.5 indicates the cut-off utilised. Droplets with bars above this cut-off were further processed, while droplets with bars below this, which have five or fewer UMIs associated to identifiable kappa chains, were rejected.

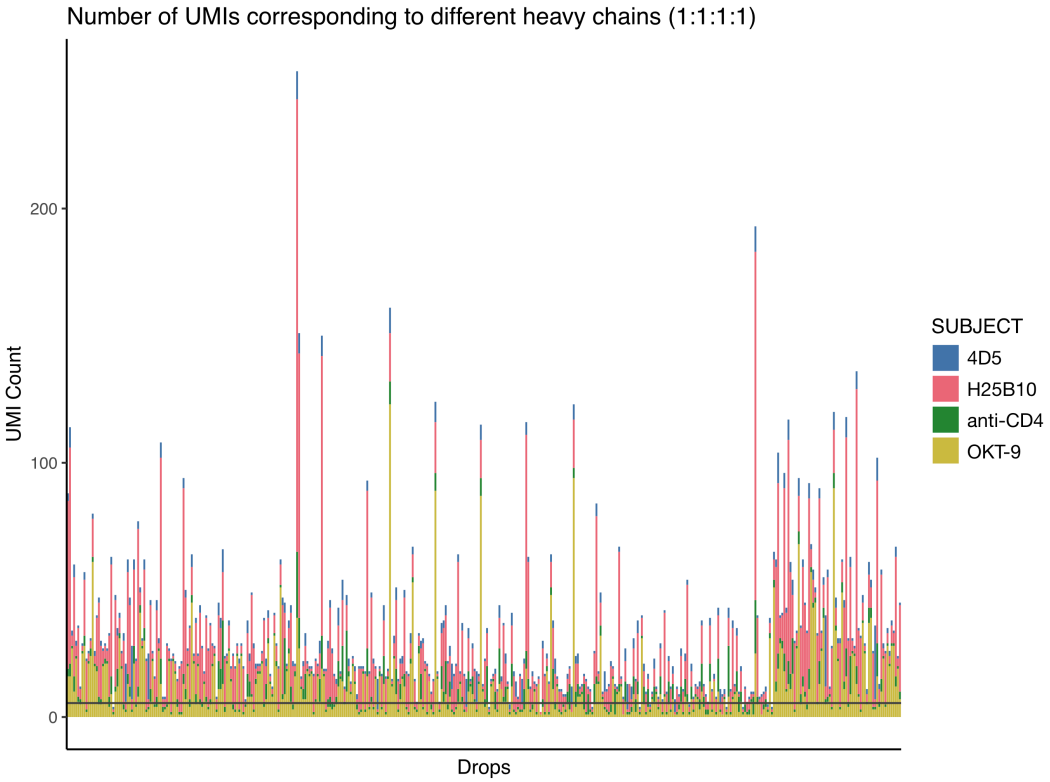


Figure E.4: Distribution of UMIs associated with droplets for heavy chain sequencing of the 1:1:1:1 hybridoma mixture. Four hybridoma cell lines (4D5, H25B10, anti-CD4 and OKT-9) were mixed in equal proportion. Independent droplets are shown on the x-axis ("Drops") while the UMI count for each droplet is depicted on the y-axis ("UMI Count"). Each bar corresponds to the UMIs associated within a single droplet. The identity of the heavy chain associated with each UMI is indicated by the colour(s) of the bar ("SUBJECT", with 4D5: blue, H25B10: pink, anti-CD4: green, OKT-9: mustard). The grey line at UMI Count = 5.5 indicates the cut-off utilised. Droplets with bars above this cut-off were further processed, while droplets with bars below this, which have five or fewer UMIs associated to identifiable heavy chains, were rejected.

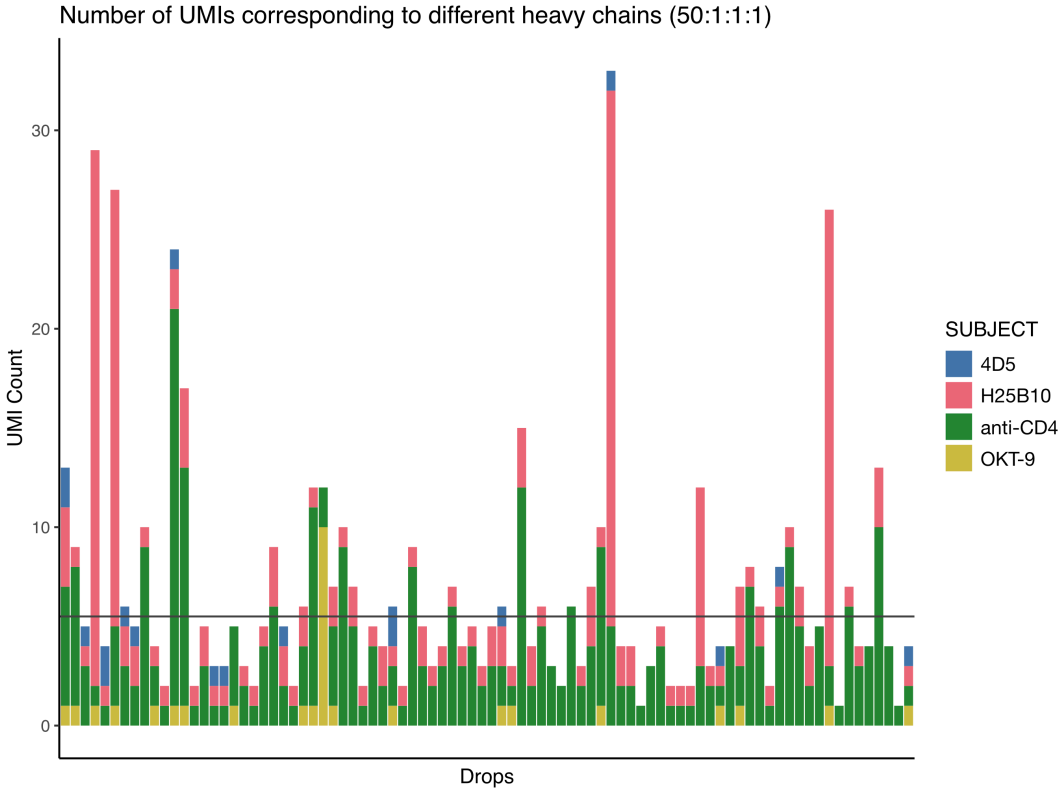


Figure E.5: Distribution of UMIs associated with droplets for heavy chain sequencing of the 50:1:1:1 hybridoma mixture. Three hybridoma cell lines (4D5, H25B10 and OKT-9) were mixed in equal proportion with anti-CD4 hybridoma cells added at 50-fold the concentration of the other cell lines. Independent droplets are shown on the x-axis ("Drops") while the UMI count for each droplet is depicted on the y-axis ("UMI Count"). Each bar corresponds to the UMIs associated within a single droplet. The identity of the heavy chain associated with each UMI is indicated by the colour(s) of the bar ("SUBJECT", with 4D5: blue, H25B10: pink, anti-CD4: green, OKT-9: mustard). The grey line at UMI Count = 5.5 indicates the cut-off utilised. Droplets with bars above this cut-off were further processed, while droplets with bars below this, which have five or fewer UMIs associated to identifiable heavy chains, were rejected.

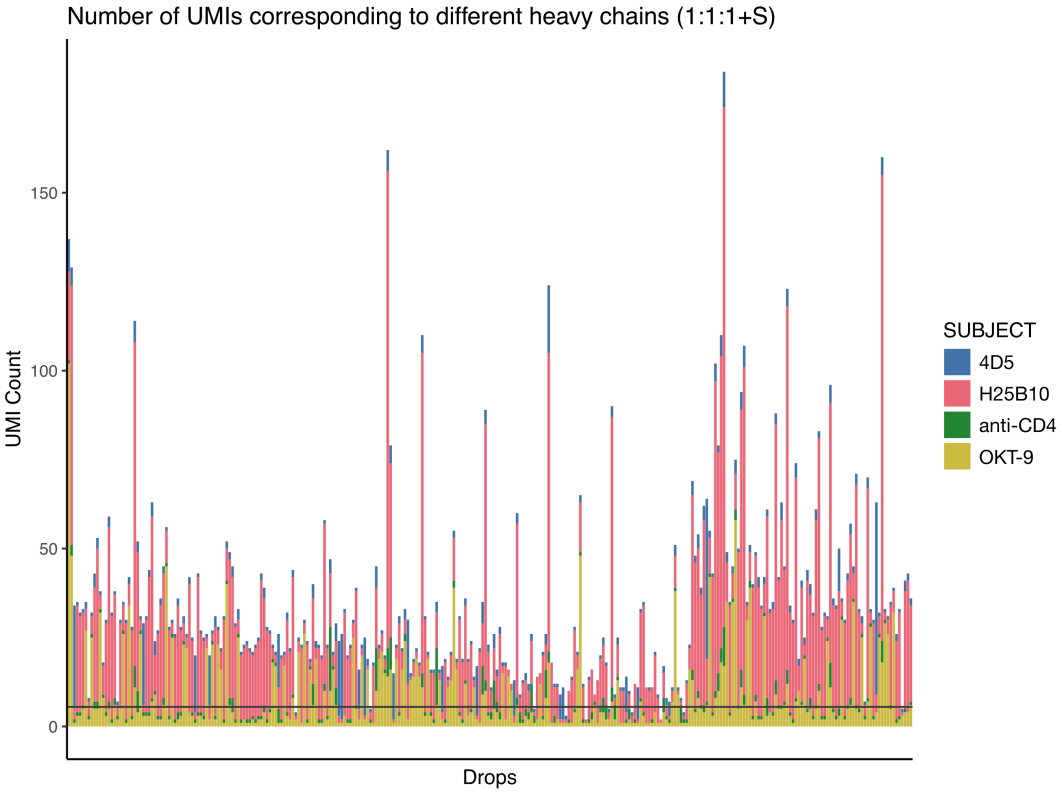


Figure E.6: Distribution of UMIs associated with droplets for heavy chain sequencing of the 1:1:1+S hybridoma mixture. Three hybridoma cell lines (4D5, H25B10 and OKT-9) were mixed in equal proportion, and incubated with anti-CD4 supernatant for 20 minutes, prior to processing. Independent droplets are shown on the x-axis ("Drops") while the UMI count for each droplet is depicted on the y-axis ("UMI Count"). Each bar corresponds to the UMIs associated within a single droplet. The identity of the heavy chain associated with each UMI is indicated by the colour(s) of the bar ("SUBJECT", with 4D5: blue, H25B10: pink, anti-CD4: green, OKT-9: mustard). The grey line at UMI Count = 5.5 indicates the cut-off utilised. Droplets with bars above this cut-off were further processed, while droplets with bars below this, which have five or fewer UMIs associated to identifiable heavy chains, were rejected.

F | Analysis of droplet code distances for unique chains

The barcodes pictured in Figure F.1 are from the sixth most frequently detected heavy chain from Group 2. The seventh and eighth barcodes (from the top) ("TCACCGACCACC" and "TCGCCGACCACC") differ by a single nucleotide in the third position (Fig. F.1). This suggests that they are unlikely to be derived from two different droplets, but are more likely to have arisen from a PCR mutation.

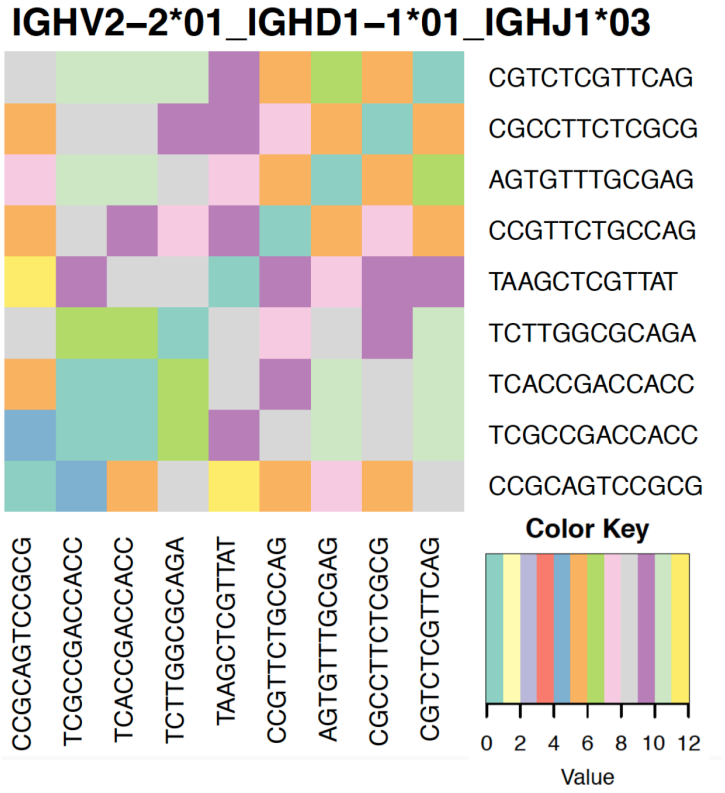


Figure F.1: Analysis of droplet code distances for the sixth most frequently detected heavy chain in Group 2. The droplet barcodes of the sixth most frequently detected heavy chain in Group 2 were examined (n = 9). Droplet barcodes are found under and to the right of the heatmap, while the different colours indicate the number of nucleotide differences between pairs of barcodes. The germline V, D, and J sequences of the antibody chain are indicated in the title. Data reproduced with the permission of Charles Girardot, EMBL Heidelberg.

G | Heavy and light chain sequences of anti-CD4 antibody

anti-CD4 heavy chain:
CAGGTCCAACTGCAGCAGCCTGGGGCTGAGCTGGTAAAGCCTGGGGCTTC
AGTGAAGTTGTCCTGCAAGGCTTCTGGCTACATTTTCACCAGTTACTGGAT
GCACTGGGTGAAGCAGAGGCCTGGACAAGGCCTTGAGTGGATTGGTATGA
TTCATCCTAATAGTGATAATACTGACTTCAATGAGAAATTCAAGAGTAAG
GTCACACTGACTGTAGACAAGTCCTCCAGCACAGCCTACATGCAACTCAG
CAGCCTGACATCTGAGGACTCTGCGGTCTATTATTGTGCAAGAGATTACTA
CGGTAGTGGCTACGGCTGGTACTTCGATGTCTGGGGCACAGGGACCACGG
TCACCGTCTCCTCAG

anti-CD4 light chain:
GACATTGTGATGACCCAGTCTCACAAGTTCATGTCCACATCAGTAGGAGA
CAGGGTCAGCATCACCTGCAAGGCCAGTCACGATGTGGGTAATGCTATAG
CCTGGTATCAACAGCGACCAGGGCGTTCTCCTAAGCTACTGATTTACTGGG
CATCCACCCGCCACACTGGAGTCCCTGATCGCTTCACAGGCAGTGGGTCT
GGGACAGATTTCTCTCTCACCATTAGCAATGTGCAGTCTGAAGACTTGGCA
GATTATTTCTGTCAGCAATATAGCAACTATCCGTTCACGTTCGGAGGGGGG
ACCAAGCTGGAAATAAAAC

Both sequences were shown to be productive when analysed with IgBLAST (Ye et al., 2013).

LaTeX

I composed this thesis with Overleaf, and used a custom document structure based on the 'Masters/Doctoral Thesis' LaTeX template (www.latextemplates.com, authors Steve Gunn, Sunil Patel, vel@latextemplates), modified by Markus Mund, which is available under CC BY-NC-SA 3.0 (http://creativecommons.org/licenses/by-nc-sa/3.0/).

Bibliography

Abate, Adam R., Tony Hung, Pascaline Mary, Jeremy J. Agresti, and David A. Weitz (Nov. 2010). "High-throughput injection with microfluidics using picoinjectors". en. In: *Proceedings of the National Academy of Sciences* 107.45, pp. 19163–19166. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1006888107.

Abidi, S. Hussain I., Tao Dong, Mai T. Vuong, Vattipally B. Sreenu, Sarah L. Rowland-Jones, Edward J. Evans, and Simon J. Davis (June 2008). "Differential remodeling of a T-cell transcriptome following CD8- versus CD3-induced signaling". In: *Cell research* 18.6, pp. 641–648. ISSN: 1001-0602. DOI: 10.1038/cr.2008.56.

Abraham, R., S. Buxbaum, J. Link, R. Smith, C. Venti, and M. Darsley (Dec. 1996). "Determination of binding constants of diabodies directed against prostate-specific antigen using electrochemiluminescence-based immunoassays". eng. In: *Journal of molecular recognition: JMR* 9.5-6, pp. 456–461. ISSN: 0952-3499. DOI: 10.1002/(sici)1099-1352(199634/12)9:5/6<456::aid-jmr282>3.0.co;2-8.

Abramowicz, Daniel, Alain Crusiaux, and Michel Goldman, M.D. (Sept. 1992). "Anaphylactic Shock after Retreatment with OKT3 Monoclonal Antibody". In: *New England Journal of Medicine* 327.10. Publisher: Massachusetts Medical Society _eprint: https://doi.org/10.1056/NEJM199209033271018, pp. 736–736. ISSN: 0028-4793. DOI: 10.1056/NEJM199209033271018.

Adler, Adam S., Rena A. Mizrahi, Matthew J. Spindler, Matthew S. Adams, Michael A. Asensio, Robert C. Edgar, Jackson Leong, Renee Leong, and David S. Johnson (Nov. 2017a). "Rare, high-affinity mouse anti-PD-1 antibodies that function in checkpoint blockade, discovered using microfluidics and molecular genomics". In: *mAbs* 9.8. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/19420862.2017.1371386, pp. 1270–1281. ISSN: 1942-0862. DOI: 10.1080/19420862.2017.1371386.

Adler, Adam S. et al. (Nov. 2017b). "Rare, high-affinity anti-pathogen antibodies from human repertoires, discovered using microfluidics and molecular genomics". In: *mAbs* 9.8. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/19420862.2017.1371383, pp. 1282–1296. ISSN: 1942-0862. DOI: 10.1080/19420862.2017.1371383.

Ahn, Keunho, Charles Kerbage, Tom P. Hunt, R. M. Westervelt, Darren R. Link, and D. A. Weitz (Jan. 2006). "Dielectrophoretic manipulation of drops for high-speed microfluidic sorting devices". In: *Applied Physics Letters* 88.2, p. 024104. ISSN: 0003-6951. DOI: 10.1063/1.2164911.

Alda-Catalinas, Celia, Danila Bredikhin, Irene Hernando-Herraez, Fátima Santos, Oana Kubinyecz, Mélanie A. Eckersley-Maslin, Oliver Stegle, and Wolf Reik (July 2020). "A Single-Cell Transcriptomics CRISPR-Activation Screen Identifies Epigenetic Regulators of the Zygotic Genome Activation Program". English. In: *Cell Systems* 11.1. Publisher: Elsevier, 25–41.e9. ISSN: 2405-4712. DOI: 10.1016/j.cels.2020.06.004.

Alewine, Christine, Raffit Hassan, and Ira Pastan (2015). "Advances in Anticancer Immunotoxin Therapy". en. In: *The Oncologist* 20.2. _eprint: https://theoncologist.onlinelibrary.wiley.com/doi/pdf/10.1634/theoncologist.2014-0358, pp. 176–185. ISSN: 1549-490X. DOI: 10.1634/theoncologist.2014-0358.

Almagro, Juan C., Tracy R. Daniels-Wells, Sonia Mayra Perez-Tapia, and Manuel L. Penichet (2017). "Progress and Challenges in the Design and Clinical Development of Antibodies for Cancer Therapy". eng. In: *Frontiers in Immunology* 8, p. 1751. ISSN: 1664-3224. DOI: 10.3389/fimmu.2017.01751.

Almanzar, Nicole et al. (July 2020). "A single-cell transcriptomic atlas characterizes ageing tissues in the mouse". en. In: *Nature*. Publisher: Nature Publishing Group, pp. 1–6. ISSN: 1476-4687. DOI: 10.1038/s41586-020-2496-1.

Amezquita, Robert A. et al. (Feb. 2020). "Orchestrating single-cell analysis with Bioconductor". en. In: *Nature Methods* 17.2. Number: 2 Publisher: Nature Publishing Group, pp. 137–145. ISSN: 1548-7105. DOI: 10.1038/s41592-019-0654-x.

Andrews, Tallulah S. and Martin Hemberg (Mar. 2019). "False signals induced by single-cell imputation". In: *F1000Research* 7. ISSN: 2046-1402. DOI: 10.12688/f1000research.16613.2.

Anna, Shelley L., Nathalie Bontoux, and Howard A. Stone (Jan. 2003). "Formation of dispersions using "flow focusing" in microchannels". en. In: *Applied Physics Letters* 82.3, pp. 364–366. ISSN: 0003-6951, 1077-3118. DOI: 10.1063/1.1537519.

Argelaguet, Ricard, Damien Arnol, Danila Bredikhin, Yonatan Deloro, Britta Velten, John C. Marioni, and Oliver Stegle (May 2020). "MOFA+: a statistical framework for comprehensive integration of multi-modal single-cell data". In: *Genome Biology* 21.1, p. 111. ISSN: 1474-760X. DOI: 10.1186/s13059-020-02015-1.

Arya, Manit, Iqbal S. Shergill, Magali Williamson, Lyndon Gommersall, Neehar Arya, and Hitendra R. H. Patel (Mar. 2005). "Basic principles of real-time quantitative PCR". eng. In: *Expert Review of Molecular Diagnostics* 5.2, pp. 209–219. ISSN: 1473-7159. DOI: 10.1586/14737159.5.2.209.

Ashkenazi, Avi and Steven M. Chamow (Oct. 1995). "Immunoadhesins: An Alternative to Human Monoclonal Antibodies". en. In: *Methods* 8.2, pp. 104–115. ISSN: 1046-2023. DOI: 10.1006/meth.1995.9996.

Awwad, Sahar and Ukrit Angkawinitwong (July 2018). "Overview of Antibody Drug Delivery". In: *Pharmaceutics* 10.3. ISSN: 1999-4923. DOI: 10.3390/pharmaceutics10030083.

Bacher, Rhonda and Christina Kendziorski (Apr. 2016). "Design and computational analysis of single-cell RNA-sequencing experiments". In: *Genome Biology* 17.1, p. 63. ISSN: 1474-760X. DOI: 10.1186/s13059-016-0927-y.

Baek, Young-Sook, Stefan Haas, Holger Hackstein, Gregor Bein, Maria Hernandez-Santana, Hans Lehrach, Sascha Sauer, and Harald Seitz (Apr. 2009). "Identification of novel transcriptional regulators involved in macrophage differentiation and activation in U937 cells". In: *BMC Immunology* 10.1, p. 18. ISSN: 1471-2172. DOI: 10.1186/1471-2172-10-18.

Barbas, C. F., A. S. Kang, R. A. Lerner, and S. J. Benkovic (Sept. 1991). "Assembly of combinatorial antibody libraries on phage surfaces: the gene III site". eng. In: *Proceedings of the National Academy of Sciences of the United States of America* 88.18, pp. 7978–7982. ISSN: 0027-8424. DOI: 10.1073/pnas.88.18.7978.

Barbosa, Eduardo José, Raimar Löbenberg, Gabriel Lima Barros de Araujo, and Nádia Araci Bou-Chacra (Aug. 2019). "Niclosamide repositioning for treating cancer: Challenges and nano-based drug delivery opportunities". en. In: *European Journal of Pharmaceutics and Biopharmaceutics* 141, pp. 58–69. ISSN: 0939-6411. DOI: 10.1016/j.ejpb.2019.05.004.

Baret, Jean-Christophe (2012). "Surfactants in droplet-based microfluidics". en. In: *Lab on a Chip* 12.3, pp. 422–433. DOI: 10.1039/C1LC20582J.

Baret, Jean-Christophe et al. (July 2009). "Fluorescence-activated droplet sorting (FADS): efficient microfluidic cell sorting based on enzymatic activity". eng. In: *Lab on a Chip* 9.13, pp. 1850–1858. ISSN: 1473-0197. DOI: 10.1039/b902504a.

Beck, Alain, Liliane Goetsch, Charles Dumontet, and Nathalie Corvaïa (May 2017). "Strategies and challenges for the next generation of antibody-drug conjugates". eng. In: *Nature Reviews. Drug Discovery* 16.5, pp. 315–337. ISSN: 1474-1784. DOI: 10.1038/nrd.2016.268.

Beer, N. Reginald et al. (Mar. 2008). "On-chip single-copy real-time reverse-transcription PCR in isolated picoliter droplets". eng. In: *Analytical Chemistry* 80.6, pp. 1854–1858. ISSN: 0003-2700. DOI: 10.1021/ac800048k.

Bent, Zachary, Elliott Meer, Daniel Riordan, Paul Ryvkin, Niranjan Srinivas, Jessica Terry, and Alex Gagnon (Feb. 2020). "Functionalized Gel Beads". Pat. 20200063191. Library Catalog: FreePatentsOnline.

Bersini, Simone, Jessie S. Jeon, Gabriele Dubini, Chiara Arrigoni, Seok Chung, Joseph L. Charest, Matteo Moretti, and Roger D. Kamm (Mar. 2014). "A microfluidic 3D in vitro model for specificity of breast cancer metastasis to bone". eng. In: *Biomaterials* 35.8, pp. 2454–2461. ISSN: 1878-5905. DOI: 10.1016/j.biomaterials.2013.11.050.

Berthier, Erwin, Jay Warrick, Hongmeiy Yu, and David J. Beebe (May 2008). "Managing evaporation for more robust microscale assays". en. In: *Lab on a Chip* 8.6, pp. 852–859. ISSN: 1473-0189. DOI: 10.1039/B717422E.

Bioč anin, Marjan, Johannes Bues, Riccardo Dainese, Esther Amstad, and Bart Deplancke (Apr. 2019). "Simplified Drop-seq workflow with minimized bead loss using a bead capture and processing microfluidic chip". en. In: *Lab on a Chip* 19.9, pp. 1610–1620. ISSN: 1473-0189. DOI: 10.1039/C9LC00014C.

Birey, Fikri et al. (May 2017). "Assembly of functionally integrated human forebrain spheroids". en. In: *Nature* 545.7652, pp. 54–59. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/nature22330.

Blondel, Vincent D., Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre (Oct. 2008). "Fast unfolding of communities in large networks". en. In: *Journal of Statistical Mechanics: Theory and Experiment* 2008.10. Publisher: IOP Publishing, P10008. ISSN: 1742-5468. DOI: 10.1088/1742-5468/2008/10/P10008.

Boder, Eric T. and K. Dane Wittrup (June 1997). "Yeast surface display for screening combinatorial polypeptide libraries". en. In: *Nature Biotechnology* 15.6. Number: 6 Publisher: Nature Publishing Group, pp. 553–557. ISSN: 1546-1696. DOI: 10.1038/nbt0697-553.

Bolger, Anthony M., Marc Lohse, and Bjoern Usadel (Aug. 2014). "Trimmomatic: a flexible trimmer for Illumina sequence data". eng. In: *Bioinformatics (Oxford, England)* 30.15, pp. 2114–2120. ISSN: 1367-4811. DOI: 10.1093/bioinformatics/btu170.

Breitling, F., S. Dübel, T. Seehaus, I. Klewinghaus, and M. Little (Aug. 1991). "A surface expression vector for antibody screening". eng. In: *Gene* 104.2, pp. 147–153. ISSN: 0378-1119. DOI: 10.1016/0378-1119(91)90244-6.

Bremond, Nicolas, Abdou R. Thiam, and Jérôme Bibette (Jan. 2008). "Decompressing Emulsion Droplets Favors Coalescence". In: *Physical Review Letters* 100.2, p. 024501. DOI: 10.1103/PhysRevLett.100.024501.

Brinkmann, Ulrich and Roland E. Kontermann (Jan. 2017). "The making of bispecific antibodies". In: *mAbs* 9.2, pp. 182–212. ISSN: 1942-0862. DOI: 10.1080/19420862.2016.1268307.

Broeder, Alfons den et al. (Nov. 2002). "A single dose, placebo controlled study of the fully human anti-tumor necrosis factor-alpha antibody adalimumab (D2E7) in patients with rheumatoid arthritis". eng. In: *The Journal of Rheumatology* 29.11, pp. 2288–2298. ISSN: 0315-162X.

Buenrostro, Jason D., M. Ryan Corces, Caleb A. Lareau, Beijing Wu, Alicia N. Schep, Martin J. Aryee, Ravindra Majeti, Howard Y. Chang, and William J. Greenleaf (May 2018). "Integrated Single-Cell Analysis Maps the Continuous Regulatory Landscape of Human Hematopoietic Differentiation". en. In: *Cell* 173.6, 1535–1548.e16. ISSN: 0092-8674. DOI: 10.1016/j.cell.2018.03.074.

Buenrostro, Jason D., Beijing Wu, Ulrike M. Litzenburger, Dave Ruff, Michael L. Gonzales, Michael P. Snyder, Howard Y. Chang, and William J. Greenleaf (July 2015). "Single-cell chromatin accessibility reveals principles of regulatory variation". en. In: *Nature* 523.7561, pp. 486–490. ISSN: 1476-4687. DOI: 10.1038/nature14590.

Busse, Christian E., Irina Czogiel, Peter Braun, Peter F. Arndt, and Hedda Wardemann (Feb. 2014). "Single-cell based high-throughput sequencing of full-length immunoglobulin heavy and light chain genes". In: *European Journal of Immunology* 44.2. Publisher: John Wiley & Sons, Ltd, pp. 597–603. ISSN: 0014-2980. DOI: 10.1002/eji.201343917.

Butler, Andrew, Paul Hoffman, Peter Smibert, Efthymia Papalexi, and Rahul Satija (June 2018). "Integrating single-cell transcriptomic data across different conditions, technologies, and species". In: *Nature biotechnology* 36.5, pp. 411–420. ISSN: 1087-0156. DOI: 10.1038/nbt.4096.

Cao, Junyue et al. (Aug. 2017). "Comprehensive single-cell transcriptional profiling of a multicellular organism". en. In: *Science* 357.6352. Publisher: American Association for the Advancement of Science Section: Research Article, pp. 661–667. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.aam8940.

Castelli, María Sofía, Paul McGonigle, and Pamela J. Hornby (Dec. 2019). "The pharmacology and therapeutic applications of monoclonal antibodies". In: *Pharmacology Research & Perspectives* 7.6. ISSN: 2052-1707. DOI: 10.1002/prp2.535.

Chabert, Max, Kevin D. Dorfman, and Jean-Louis Viovy (2005). "Droplet fusion by alternating current (AC) field electrocoalescence in microchannels". en. In: *ELECTROPHORESIS* 26.19, pp. 3706–3715. ISSN: 1522-2683. DOI: 10.1002/elps.200500109.

Chan, C. J., M. J. Smyth, and L. Martinet (Jan. 2014a). "Molecular mechanisms of natural killer cell activation in response to cellular stress". en. In: *Cell Death & Differentiation* 21.1, pp. 5–14. ISSN: 1350-9047. DOI: 10.1038/cdd.2013.26.

Chan, Conrad E. Z., Annie H. Y. Chan, Angeline P. C. Lim, and Brendon J. Hanson (Oct. 2011). "Comparison of the efficiency of antibody selection from semi-synthetic scFv and non-immune Fab phage display libraries against protein targets for rapid development of diagnostic immunoassays". eng. In: *Journal of Immunological Methods* 373.1-2, pp. 79–88. ISSN: 1872-7905. DOI: 10.1016/j.jim.2011.08.005.

Chan, Kin Tak, Mei Yuk Choi, Kenneth K. Y. Lai, Winnie Tan, Lai Nar Tung, Ho Yu Lam, Daniel K. H. Tong, Nikki P. Lee, and Simon Law (Mar. 2014b). "Overexpression of transferrin receptor CD71 and its tumorigenic properties in esophageal squamous cell carcinoma". In: *Oncology Reports* 31.3. Publisher: Spandidos Publications, pp. 1296–1304. ISSN: 1021-335X. DOI: 10.3892/or.2014.2981.

Charles A Janeway, Jr, Paul Travers, Mark Walport, and Mark J. Shlomchik (2001). "Antigen Recognition by B-cell and T-cell Receptors". en. In: *Immunobiology: The Immune System in Health and Disease. 5th edition*. Publisher: Garland Science.

Chen, Jin-Shing, Kengli Lan, and Mien-Chie Hung (June 2003). "Strategies to target HER2/neu overexpression for cancer therapy". eng. In: *Drug Resistance Updates: Reviews and Commentaries in Antimicrobial and Anticancer Chemotherapy* 6.3, pp. 129–136. ISSN: 1368-7646. DOI: 10.1016/s1368-7646(03)00040-2.

Chen, Wei, Robert A. Mook, Richard T. Premont, and Jiangbo Wang (Jan. 2018). "Niclosamide: Beyond an antihelminthic drug". In: *Cellular Signalling* 41, pp. 89–96. ISSN: 0898-6568. DOI: 10.1016/j.cellsig.2017.04.001.

Chokshi, Trushal Vijaykumar, Adela Ben-Yakar, and Nikos Chronis (Jan. 2009). "CO2 and compressive immobilization of C. elegans on-chip". eng. In: *Lab on a Chip* 9.1, pp. 151–157. ISSN: 1473-0197. DOI: 10.1039/b807345g.

Choudhury, Deepak, Danny van Noort, Ciprian Iliescu, Baixue Zheng, Kar-Lai Poon, Svetlana Korzh, Vladimir Korzh, and Hanry Yu (Feb. 2012). "Fish and Chips: a microfluidic perfusion platform for monitoring zebrafish development". en. In: *Lab on a Chip* 12.5, pp. 892–900. ISSN: 1473-0189. DOI: 10.1039/C1LC20351G.

Chowdhury, Mohammad Suman, Wenshan Zheng, Shalini Kumari, John Heyman, Xingcai Zhang, Pradip Dey, David A. Weitz, and Rainer Haag (Oct. 2019). "Dendronized fluorosurfactant for highly stable water-in-fluorinated oil emulsions with minimal inter-droplet transfer of small molecules". en. In: *Nature Communications* 10.1, pp. 1–10. ISSN: 2041-1723. DOI: 10.1038/s41467-019-12462-5.

Chung, Kwanghun, Yoosik Kim, Jitendra S. Kanodia, Emily Gong, Stanislav Y. Shvartsman, and Hang Lu (Feb. 2011). "A microfluidic array for large-scale ordering and orientation of embryos". In: *Nature methods* 8.2, pp. 171–176. ISSN: 1548-7091. DOI: 10.1038/nmeth.1548.

Clausell-Tormos, Jenifer et al. (May 2008). "Droplet-based microfluidic platforms for the encapsulation and screening of Mammalian cells and multicellular organisms". eng. In:

Chemistry & Biology 15.5, pp. 427–437. ISSN: 1074-5521. DOI: 10.1016/j.chembiol.2008.04.004.

Convery, Neil and Nikolaj Gadegaard (Mar. 2019). "30 years of microfluidics". en. In: *Micro and Nano Engineering* 2, pp. 76–91. ISSN: 2590-0072. DOI: 10.1016/j.mne.2019.01.003.

Cronin, James, Xian-Yang Zhang, and Jakob Reiser (Aug. 2005). "Altering the Tropism of Lentiviral Vectors through Pseudotyping". In: *Current gene therapy* 5.4, pp. 387–398. ISSN: 1566-5232.

Cusanovich, Darren A., Riza Daza, Andrew Adey, Hannah A. Pliner, Lena Christiansen, Kevin L. Gunderson, Frank J. Steemers, Cole Trapnell, and Jay Shendure (May 2015). "Multiplex single-cell profiling of chromatin accessibility by combinatorial cellular indexing". en. In: *Science* 348.6237. Publisher: American Association for the Advancement of Science Section: Report, pp. 910–914. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.aab1601.

Daniels, Tracy R. et al. (Mar. 2012). "Transferrin receptors and the targeted delivery of therapeutic agents against cancer". In: *Biochimica et biophysica acta* 1820.3, pp. 291–317. ISSN: 0006-3002. DOI: 10.1016/j.bbagen.2011.07.016.

Datlinger, Paul et al. (2017). "Pooled CRISPR screening with single-cell transcriptome readout". eng. In: *Nature Methods* 14.3, pp. 297–301. ISSN: 1548-7105. DOI: 10.1038/nmeth.4177.

Dealmakers, BioPharma (Sept. 2019). *Moving up with the monoclonals*. en. Library Catalog: biopharmadealmakers.nature.com Section: Antibodies.

DeKosky, Brandon J., Takaaki Kojima, Alexa Rodin, Wissam Charab, Gregory C. Ippolito, Andrew D. Ellington, and George Georgiou (Jan. 2015). "In-depth determination and analysis of the human paired heavy- and light-chain antibody repertoire". en. In: *Nature Medicine* 21.1. Number: 1 Publisher: Nature Publishing Group, pp. 86–91. ISSN: 1546-170X. DOI: 10.1038/nm.3743.

DeKosky, Brandon J et al. (Feb. 2013). "High-throughput sequencing of the paired human immunoglobulin heavy and light chain repertoire". In: *Nature biotechnology* 31.2, pp. 166–169. ISSN: 1087-0156. DOI: 10.1038/nbt.2492.

Dey, Siddharth S., Lennart Kester, Bastiaan Spanjaard, Magda Bienko, and Alexander van Oudenaarden (Mar. 2015). "Integrated genome and transcriptome sequencing of the same cell". eng. In: *Nature Biotechnology* 33.3, pp. 285–289. ISSN: 1546-1696. DOI: 10.1038/nbt.3129.

Dijk, David van et al. (July 2018). "Recovering gene interactions from single-cell data using data diffusion". In: *Cell* 174.3, 716–729.e27. ISSN: 0092-8674. DOI: 10.1016/j.cell.2018.05.061.

Ding, Jiarui et al. (May 2019). "Systematic comparative analysis of single cell RNA-sequencing methods". en. In: *bioRxiv*. Publisher: Cold Spring Harbor Laboratory Section: New Results, p. 632216. DOI: 10.1101/632216.

Dixit, Atray et al. (Dec. 2016). "Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens". en. In: *Cell* 167.7, 1853–1866.e17. ISSN: 0092-8674. DOI: 10.1016/j.cell.2016.11.038.

Dobin, Alexander, Carrie A. Davis, Felix Schlesinger, Jorg Drenkow, Chris Zaleski, Sonali Jha, Philippe Batut, Mark Chaisson, and Thomas R. Gingeras (Jan. 2013). "STAR: ultrafast universal RNA-seq aligner". en. In: *Bioinformatics* 29.1. Publisher: Oxford Academic, pp. 15–21. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/bts635.

Dooley, Helen, Martin F. Flajnik, and Andrew J. Porter (Sept. 2003). "Selection and characterization of naturally occurring single-domain (IgNAR) antibody fragments from immunized sharks by phage display". eng. In: *Molecular Immunology* 40.1, pp. 25–33. ISSN: 0161-5890. DOI: 10.1016/s0161-5890(03)00084-1.

Du, Wenbin, Liang Li, Kevin P. Nichols, and Rustem F. Ismagilov (Aug. 2009). "SlipChip". eng. In: *Lab on a Chip* 9.16, pp. 2286–2292. ISSN: 1473-0197. DOI: 10.1039/b908978k.

Duan, Bin et al. (May 2019). "Model-based understanding of single-cell CRISPR screening". en. In: *Nature Communications* 10.1. Number: 1 Publisher: Nature Publishing Group, p. 2233. ISSN: 2041-1723. DOI: 10.1038/s41467-019-10216-x.

Duffy, David C., J. Cooper McDonald, Olivier J. A. Schueller, and George M. Whitesides (Dec. 1998). "Rapid Prototyping of Microfluidic Systems in Poly(dimethylsiloxane)". In: *Analytical Chemistry* 70.23, pp. 4974–4984. ISSN: 0003-2700. DOI: 10.1021/ac980656z.

Dura, Burak, Stephanie K. Dougan, Marta Barisa, Melanie M. Hoehl, Catherine T. Lo, Hidde L. Ploegh, and Joel Voldman (Jan. 2015). "Profiling lymphocyte interactions at the single-cell level by microfluidic cell pairing". eng. In: *Nature Communications* 6, p. 5940. ISSN: 2041-1723. DOI: 10.1038/ncomms6940.

Dura, Burak, Mariah M. Servos, Rachel M. Barry, Hidde L. Ploegh, Stephanie K. Dougan, and Joel Voldman (2016). "Longitudinal multiparameter assay of lymphocyte interactions from onset by microfluidic cell pairing and culture". eng. In: *Proceedings of the National Academy of Sciences of the United States of America* 113.26, E3599–3608. ISSN: 1091-6490. DOI: 10.1073/pnas.1515364113.

Eastburn, Dennis J., Adam Sciambi, and Adam R. Abate (Apr. 2013a). "Picoinjection Enables Digital Detection of RNA with Droplet RT-PCR". en. In: *PLOS ONE* 8.4. Publisher: Public Library of Science, e62961. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0062961.

— (Aug. 2013b). "Ultrahigh-Throughput Mammalian Single-Cell Reverse-Transcriptase Polymerase Chain Reaction in Microfluidic Drops". In: *Analytical Chemistry* 85.16, pp. 8016–8021. ISSN: 0003-2700. DOI: 10.1021/ac402057q.

Eduati, Federica, Ramesh Utharala, Dharanija Madhavan, Ulf Peter Neumann, Thomas Longerich, Thorsten Cramer, Julio Saez-Rodriguez, and Christoph A. Merten (June 2018). "A microfluidics platform for combinatorial drug screening on cancer biopsies". en. In: *Nature Communications* 9.1, pp. 1–13. ISSN: 2041-1723. DOI: 10.1038/s41467-018-04919-w.

Eisen, Herman N. (May 2014). "Affinity enhancement of antibodies: how low-affinity antibodies produced early in immune responses are followed by high-affinity antibodies later and in memory B-cell responses". eng. In: *Cancer Immunology Research* 2.5, pp. 381–392. ISSN: 2326-6074. DOI: 10.1158/2326-6066.CIR-14-0029.

El Debs, Bachir, Ramesh Utharala, Irina V. Balyasnikova, Andrew D. Griffiths, and Christoph A. Merten (July 2012). "Functional single-cell hybridoma screening using droplet-based microfluidics". eng. In: *Proceedings of the National Academy of Sciences of the United States of America* 109.29, pp. 11570–11575. ISSN: 1091-6490. DOI: 10.1073/pnas.1204514109.

Elhanati, Yuval, Zachary Sethna, Quentin Marcou, Curtis G. Callan, Thierry Mora, and Aleksandra M. Walczak (Sept. 2015). "Inferring processes underlying B-cell repertoire diversity". eng. In: *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 370.1676. ISSN: 1471-2970. DOI: 10.1098/rstb.2014.0243.

Eraslan, Gökcen, Lukas M. Simon, Maria Mircea, Nikola S. Mueller, and Fabian J. Theis (Jan. 2019). "Single-cell RNA-seq denoising using a deep count autoencoder". en. In: *Nature Communications* 10.1. Number: 1 Publisher: Nature Publishing Group, p. 390. ISSN: 2041-1723. DOI: 10.1038/s41467-018-07931-2.

Eyer, Klaus et al. (Oct. 2017). "Single-cell deep phenotyping of IgG-secreting cells for high-resolution immune monitoring". eng. In: *Nature Biotechnology* 35.10, pp. 977–982. ISSN: 1546-1696. DOI: 10.1038/nbt.3964.

Fabi, Alessandra, Paola Malaguti, Sabrina Vari, and Francesco Cognetti (June 2016). "First-line therapy in HER2 positive metastatic breast cancer: is the mosaic fully completed or are we missing additional pieces?" eng. In: *Journal of experimental & clinical cancer research: CR* 35, p. 104. ISSN: 1756-9966. DOI: 10.1186/s13046-016-0380-5.

Fan, H. Christina, Glenn K. Fu, and Stephen P. A. Fodor (Feb. 2015). "Combinatorial labeling of single cells for gene expression cytometry". en. In: *Science* 347.6222. Publisher: American Association for the Advancement of Science Section: Research Article. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.1258367.

Fan, H Christina, Jianbin Wang, Anastasia Potanina, and Stephen R Quake (Jan. 2011). "Whole-genome molecular haplotyping of single cells". In: *Nature biotechnology* 29.1, pp. 51–57. ISSN: 1087-0156. DOI: 10.1038/nbt.1739.

Finak, Greg et al. (Dec. 2015). "MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data". In: *Genome Biology* 16.1, p. 278. ISSN: 1474-760X. DOI: 10.1186/s13059-015-0844-5.

Finn, Richard S. et al. (Oct. 2009). "PD 0332991, a selective cyclin D kinase 4/6 inhibitor, preferentially inhibits proliferation of luminal estrogen receptor-positive human breast cancer cell lines in vitro". In: *Breast Cancer Research* 11.5, R77. ISSN: 1465-542X. DOI: 10.1186/bcr2419.

Frenzel, Daniel and Christoph A. Merten (Mar. 2017). "Microfluidic train station: highly robust and multiplexable sorting of droplets on electric rails". en. In: *Lab on a Chip* 17.6. Publisher: The Royal Society of Chemistry, pp. 1024–1030. ISSN: 1473-0189. DOI: 10.1039/C6LC01544A.

Georgiou, George, Gregory C Ippolito, John Beausang, Christian E Busse, Hedda Wardemann, and Stephen R Quake (2014). "The promise and challenge of high-throughput sequencing of the antibody repertoire". In: *Nature Biotechnology* 32.2, pp. 158–168. ISSN: 1087-0156. DOI: 10.1038/nbt.2782.

Giaever, I. and C. R. Keese (Jan. 1983). "Behavior of cells at fluid interfaces". en. In: *Proceedings of the National Academy of Sciences* 80.1, pp. 219–222. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.80.1.219.

Gierahn, Todd M., Marc H. Wadsworth, Travis K. Hughes, Bryan D. Bryson, Andrew Butler, Rahul Satija, Sarah Fortune, J. Christopher Love, and Alex K. Shalek (Apr. 2017). "Seq-Well: portable, low-cost RNA sequencing of single cells at high throughput". eng. In: *Nature Methods* 14.4, pp. 395–398. ISSN: 1548-7105. DOI: 10.1038/nmeth.4179.

Girardot, Charles, Jelle Scholtalbers, Sajoscha Sauer, Shu-Yi Su, and Eileen E.M. Furlong (Oct. 2016). "Je, a versatile suite to handle multiplexed NGS libraries with unique molecular identifiers". In: *BMC Bioinformatics* 17.1, p. 419. ISSN: 1471-2105. DOI: 10.1186/s12859-016-1284-2.

Glatzová, Daniela and Marek Cebecauer (Apr. 2019). "Dual Role of CD4 in Peripheral T Lymphocytes". In: *Frontiers in Immunology* 10. ISSN: 1664-3224. DOI: 10.3389/fimmu.2019.00618.

Goers, Lisa, Paul Freemont, and Karen M. Polizzi (July 2014). "Co-culture systems and technologies: taking synthetic biology to the next level". In: *Journal of The Royal Society Interface* 11.96. Publisher: Royal Society, p. 20140065. DOI: 10.1098/rsif.2014.0065.

Goldstein, Leonard D. et al. (July 2017). "Massively parallel nanowell-based single-cell gene expression profiling". In: *BMC Genomics* 18. ISSN: 1471-2164. DOI: 10.1186/s12864-017-3893-1.

Goldstein, Leonard D. et al. (Aug. 2019). "Massively parallel single-cell B-cell receptor sequencing enables rapid discovery of diverse antigen-reactive antibodies". en. In: *Communications Biology* 2.1. Number: 1 Publisher: Nature Publishing Group, pp. 1–10. ISSN: 2399-3642. DOI: 10.1038/s42003-019-0551-y.

Guo, Feng, Jarrod B. French, Peng Li, Hong Zhao, Chung Yu Chan, James R. Fick, Stephen J. Benkovic, and Tony Jun Huang (Aug. 2013). "Probing cell-cell communication with microfluidic devices". In: *Lab on a chip* 13.16, pp. 3152–3162. ISSN: 1473-0197. DOI: 10.1039/c3lc90067c.

Gérard, Annabelle et al. (June 2020). "High-throughput single-cell activity-based screening and sequencing of antibodies using droplet microfluidics". en. In: *Nature Biotechnology* 38.6, pp. 715–721. ISSN: 1087-0156, 1546-1696. DOI: 10.1038/s41587-020-0466-7.

Hall, Neil (May 2013). "After the gold rush". In: *Genome Biology* 14.5, p. 115. ISSN: 1474-760X. DOI: 10.1186/gb-2013-14-5-115.

Han, Qing, Neda Bagheri, Elizabeth M. Bradshaw, David A. Hafler, Douglas A. Lauffenburger, and J. Christopher Love (Jan. 2012). "Polyfunctional responses by

human T cells result from sequential release of cytokines". en. In: *Proceedings of the National Academy of Sciences* 109.5, pp. 1607–1612. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1117194109.

Hansel, Trevor T., Harald Kropshofer, Thomas Singer, Jane A. Mitchell, and Andrew J. T. George (Apr. 2010). "The safety and side effects of monoclonal antibodies". en. In: *Nature Reviews Drug Discovery* 9.4. Number: 4 Publisher: Nature Publishing Group, pp. 325–338. ISSN: 1474-1784. DOI: 10.1038/nrd3003.

Haque, Ashraful, Jessica Engel, Sarah A. Teichmann, and Tapio Lönnberg (Aug. 2017). "A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications". In: *Genome Medicine* 9.1, p. 75. ISSN: 1756-994X. DOI: 10.1186/s13073-017-0467-4.

Hashimshony, Tamar, Florian Wagner, Noa Sher, and Itai Yanai (Sept. 2012). "CEL-Seq: single-cell RNA-Seq by multiplexed linear amplification". eng. In: *Cell Reports* 2.3, pp. 666–673. ISSN: 2211-1247. DOI: 10.1016/j.celrep.2012.08.003.

Henjes, F. et al. (July 2012). "Strong EGFR signaling in cell line models of ERBB2-amplified breast cancer attenuates response towards ERBB2-targeting drugs". en. In: *Oncogenesis* 1.7, e16. DOI: 10.1038/oncsis.2012.16.

Heyries, Kevin A., Carolina Tropini, Michael VanInsberghe, Callum Doolin, Oleh I. Petriv, Anupam Singhal, Kaston Leung, Curtis B. Hughesman, and Carl L. Hansen (Aug. 2011). "Megapixel digital PCR". en. In: *Nature Methods* 8.8. Number: 8 Publisher: Nature Publishing Group, pp. 649–651. ISSN: 1548-7105. DOI: 10.1038/nmeth.1640.

Hicks, Stephanie C., F. William Townes, Mingxiang Teng, and Rafael A. Irizarry (Oct. 2018). "Missing data and technical variability in single-cell RNA-sequencing experiments". en. In: *Biostatistics* 19.4. Publisher: Oxford Academic, pp. 562–578. ISSN: 1465-4644. DOI: 10.1093/biostatistics/kxx053.

Hill, Andrew J., José L. McFaline-Figueroa, Lea M. Starita, Molly J. Gasperini, Kenneth A. Matreyek, Jonathan Packer, Dana Jackson, Jay Shendure, and Cole Trapnell (2018). "On the design of CRISPR-based single-cell molecular screens". eng. In: *Nature Methods* 15.4, pp. 271–274. ISSN: 1548-7105. DOI: 10.1038/nmeth.4604.

Hoehn, Kenneth B., Anna Fowler, Gerton Lunter, and Oliver G. Pybus (May 2016). "The Diversity and Molecular Evolution of B-Cell Receptors during Infection". In: *Molecular Biology and Evolution* 33.5, pp. 1147–1157. ISSN: 0737-4038. DOI: 10.1093/molbev/msw015.

Hoffman, William, Fadi G. Lakkis, and Geetha Chalasani (Jan. 2016). "B Cells, Antibodies, and More". In: *Clinical Journal of the American Society of Nephrology : CJASN* 11.1, pp. 137–154. ISSN: 1555-9041. DOI: 10.2215/CJN.09430915.

Holtze, C. et al. (Oct. 2008). "Biocompatible surfactants for water-in-fluorocarbon emulsions". eng. In: *Lab on a Chip* 8.10, pp. 1632–1639. ISSN: 1473-0197. DOI: 10.1039/b806706f.

Hosseini, Samira, Patricia Vázquez-Villegas, and Sergio O. Martínez-Chapa (Aug. 2017). "Paper and Fiber-Based Bio-Diagnostic Platforms: Current Challenges and Future Needs". en. In: *Applied Sciences* 7.8, p. 863. DOI: 10.3390/app7080863.

Hu, Hongxing, David Eustace, and Christoph A. Merten (Oct. 2015). "Efficient cell pairing in droplets using dual-color sorting". eng. In: *Lab on a Chip* 15.20, pp. 3989–3993. ISSN: 1473-0189. DOI: 10.1039/c5lc00686d.

Huang, Mo et al. (July 2018). "SAVER: Gene expression recovery for single-cell RNA sequencing". In: *Nature methods* 15.7, pp. 539–542. ISSN: 1548-7091. DOI: 10.1038/s41592-018-0033-z.

Huberts, Daphne H. E. W., Sung Sik Lee, Javier Gonzáles, Georges E. Janssens, Ima Avalos Vizcarra, and Matthias Heinemann (June 2013). "Construction and use of a microfluidic dissection platform for long-term imaging of cellular processes in budding yeast". eng. In: *Nature Protocols* 8.6, pp. 1019–1027. ISSN: 1750-2799. DOI: 10.1038/nprot.2013.060.

Hwang, Byungjin, Ji Hyun Lee, and Duhee Bang (Aug. 2018). "Single-cell RNA sequencing technologies and bioinformatics pipelines". en. In: *Experimental & Molecular Medicine*

50.8. Number: 8 Publisher: Nature Publishing Group, pp. 1–14. ISSN: 2092-6413. DOI: 10.1038/s12276-018-0071-8.

Islam, Saiful, Amit Zeisel, Simon Joost, Gioele La Manno, Pawel Zajac, Maria Kasper, Peter Lönnerberg, and Sten Linnarsson (Feb. 2014). "Quantitative single-cell RNA-seq with unique molecular identifiers". en. In: *Nature Methods* 11.2, pp. 163–166. ISSN: 1548-7105. DOI: 10.1038/nmeth.2772.

Jaffers, GJ, TC Fuller, Cosimi AB, Russell PS, Winn HJ, and Colvin RB (1986). "Monoclonal antibody therapy. Anti-idiotypic and non-anti-idiotypic antibodies to OKT3 arising despite intense immunosuppression." In:

Jain, Namrata G., Elisabeth A. Wong, Alexander J. Aranyosi, Leo Boneschansker, James F. Markmann, David M. Briscoe, and Daniel Irimia (Nov. 2015). "Microfluidic mazes to characterize T-cell exploration patterns following activation in vitro". eng. In: *Integrative Biology: Quantitative Biosciences from Nano to Macro* 7.11, pp. 1423–1431. ISSN: 1757-9708. DOI: 10.1039/c5ib00146c.

Jain, Rakesh K., Gerald C. Koenig, Marc Dellian, Dai Fukumura, Lance L. Munn, and Robert J. Melder (June 1996). "Leukocyte-endothelial adhesion and angiogenesis in tumors". en. In: *Cancer and Metastasis Reviews* 15.2, pp. 195–204. ISSN: 0167-7659, 1573-7233. DOI: 10.1007/BF00437472.

Jin, Aishun et al. (Sept. 2009). "A rapid and efficient single-cell manipulation method for screening antigen-specific antibody–secreting cells from human peripheral blood". en. In: *Nature Medicine* 15.9. Number: 9 Publisher: Nature Publishing Group, pp. 1088–1092. ISSN: 1546-170X. DOI: 10.1038/nm.1966.

Jones, M.D. (2010). *Monoclonals.png (PNG Image, 1308 × 1464 pixels)*.

Junkin, Michael, Alicia J. Kaestli, Zhang Cheng, Christian Jordi, Cem Albayrak, Alexander Hoffmann, and Savaş Tay (Apr. 2016). "High-Content Quantification of Single-Cell Immune Dynamics". eng. In: *Cell Reports* 15.2, pp. 411–422. ISSN: 2211-1247. DOI: 10.1016/j.celrep.2016.03.033.

Kamath, Amrita V. (Sept. 2016). "Translational pharmacokinetics and pharmacodynamics of monoclonal antibodies". en. In: *Drug Discovery Today: Technologies*. Technology – Translational Pharmacology 21-22, pp. 75–83. ISSN: 1740-6749. DOI: 10.1016/j.ddtec.2016.09.004.

Kaminski, Denise A., Chungwen Wei, Yu Qian, Alexander F. Rosenberg, and Ignacio Sanz (2012). "Advances in Human B Cell Phenotypic Profiling". English. In: *Frontiers in Immunology* 3. Publisher: Frontiers. ISSN: 1664-3224. DOI: 10.3389/fimmu.2012.00302.

Karzbrun, Eyal, Aditya Kshirsagar, Sidney R. Cohen, Jacob H. Hanna, and Orly Reiner (May 2018). "Human Brain Organoids on a Chip Reveal the Physics of Folding". In: *Nature physics* 14.5, pp. 515–522. ISSN: 1745-2473. DOI: 10.1038/s41567-018-0046-7.

Kashima, Yukie, Ayako Suzuki, and Yutaka Suzuki (2019). "An Informative Approach to Single-Cell Sequencing Analysis". eng. In: *Advances in Experimental Medicine and Biology* 1129, pp. 81–96. ISSN: 0065-2598. DOI: 10.1007/978-981-13-6037-4_6.

Kebschull, Justus M. and Anthony M. Zador (Nov. 2018). "Cellular barcoding: lineage tracing, screening and beyond". en. In: *Nature Methods* 15.11. Number: 11 Publisher: Nature Publishing Group, pp. 871–879. ISSN: 1548-7105. DOI: 10.1038/s41592-018-0185-x.

Kemna, Evelien W. M., Rogier M. Schoeman, Floor Wolbers, Istvan Vermes, David A. Weitz, and Albert van den Berg (July 2012). "High-yield cell ordering and deterministic cell-in-droplet encapsulation using Dean flow in a curved microchannel". en. In: *Lab on a Chip* 12.16. Publisher: The Royal Society of Chemistry, pp. 2881–2887. ISSN: 1473-0189. DOI: 10.1039/C2LC00013J.

Kharchenko, Peter V., Lev Silberstein, and David T. Scadden (July 2014). "Bayesian approach to single-cell differential expression analysis". en. In: *Nature Methods* 11.7. Number: 7 Publisher: Nature Publishing Group, pp. 740–742. ISSN: 1548-7105. DOI: 10.1038/nmeth.2967.

Kim, Choong, Jae Hoon Bang, Young Eun Kim, Soo Hyun Lee, and Ji Yoon Kang (Oct. 2012). "On-chip anticancer drug test of regular tumor spheroids formed in microwells by a distributive microchannel network". eng. In: *Lab on a Chip* 12.20, pp. 4135–4142. ISSN: 1473-0189. DOI: 10.1039/c2lc40570a.

Kim, Jonghan, William L. Hayton, John M. Robinson, and Clark L. Anderson (Feb. 2007). "Kinetics of FcRn-mediated recycling of IgG and albumin in human: Pathophysiology and therapeutic implications using a simplified mechanism-based model". In: *Clinical immunology (Orlando, Fla.)* 122.2, pp. 146–155. ISSN: 1521-6616. DOI: 10.1016/j.clim.2006.09.001.

Kim, Sangbae et al. (May 2019). "Generation, transcriptome profiling, and functional validation of cone-rich human retinal organoids". en. In: *Proceedings of the National Academy of Sciences* 116.22, pp. 10824–10833. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1901572116.

Kiselev, Vladimir Yu, Tallulah S. Andrews, and Martin Hemberg (May 2019). "Challenges in unsupervised clustering of single-cell RNA-seq data". en. In: *Nature Reviews Genetics* 20.5. Number: 5 Publisher: Nature Publishing Group, pp. 273–282. ISSN: 1471-0064. DOI: 10.1038/s41576-018-0088-9.

Klammt, Christian, Lucie Novotná, Dongyang T. Li, Miriam Wolf, Amy Blount, Kai Zhang, Jonathan R. Fitchett, and Björn F. Lillemeier (Sept. 2015). "T cell receptor dwell times control the kinase activity of Zap70". en. In: *Nature Immunology* 16.9. Number: 9 Publisher: Nature Publishing Group, pp. 961–969. ISSN: 1529-2916. DOI: 10.1038/ni.3231.

Klein, Allon M, Linas Mazutis, Ilke Akartuna, Naren Tallapragada, Adrian Veres, Victor Li, Leonid Peshkin, David A Weitz, and Marc W Kirschner (May 2015). "Droplet barcoding for single cell transcriptomics applied to embryonic stem cells". In: *Cell* 161.5, pp. 1187–1201. ISSN: 0092-8674. DOI: 10.1016/j.cell.2015.04.044.

Knappik, A. et al. (Feb. 2000). "Fully synthetic human combinatorial antibody libraries (HuCAL) based on modular consensus frameworks and CDRs randomized with trinucleotides". eng. In: *Journal of Molecular Biology* 296.1, pp. 57–86. ISSN: 0022-2836. DOI: 10.1006/jmbi.1999.3444.

Kolodziejczyk, Aleksandra A., Jong Kyoung Kim, Valentine Svensson, John C. Marioni, and Sarah A. Teichmann (May 2015). "The Technology and Biology of Single-Cell RNA Sequencing". en. In: *Molecular Cell* 58.4, pp. 610–620. ISSN: 1097-2765. DOI: 10.1016/j.molcel.2015.04.005.

Kontermann, Roland E. (Dec. 2011). "Strategies for extended serum half-life of protein therapeutics". eng. In: *Current Opinion in Biotechnology* 22.6, pp. 868–876. ISSN: 1879-0429. DOI: 10.1016/j.copbio.2011.06.012.

Krüger, Carina et al. (July 2002). "In situ delivery of passive immunity by lactobacilli producing single-chain antibodies". en. In: *Nature Biotechnology* 20.7. Number: 7 Publisher: Nature Publishing Group, pp. 702–706. ISSN: 1546-1696. DOI: 10.1038/nbt0702-702.

Kwakkenbos, Mark J. et al. (Jan. 2010). "Generation of stable monoclonal antibody-producing B cell receptor-positive human memory B cells by genetic programming". eng. In: *Nature Medicine* 16.1, pp. 123–128. ISSN: 1546-170X. DOI: 10.1038/nm.2071.

Kwong, Peter D., Richard Wyatt, James Robinson, Raymond W. Sweet, Joseph Sodroski, and Wayne A. Hendrickson (June 1998). "Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody". In: *Nature* 393.6686, pp. 648–659. ISSN: 0028-0836. DOI: 10.1038/31405.

Köhler, G. and C. Milstein (Aug. 1975). "Continuous cultures of fused cells secreting antibody of predefined specificity". en. In: *Nature* 256.5517, pp. 495–497. ISSN: 1476-4687. DOI: 10.1038/256495a0.

Labrijn, Aran F., Maarten L. Janmaat, Janice M. Reichert, and Paul W. H. I. Parren (Aug. 2019). "Bispecific antibodies: a mechanistic review of the pipeline". en. In: *Nature Reviews Drug Discovery* 18.8. Number: 8 Publisher: Nature Publishing Group, pp. 585–608. ISSN: 1474-1784. DOI: 10.1038/s41573-019-0028-1.

Lamberti, A., S. L. Marasso, and M. Cocuzza (2014). "PDMS membranes with tunable gas permeability for microfluidic applications". en. In: *RSC Advances* 4.106, pp. 61415–61419. DOI: 10.1039/C4RA12934B.

Lan, Freeman, Benjamin Demaree, Noorsher Ahmed, and A Abate (July 2017). "SiC-Seq: Single-cell genome sequencing at ultra high-throughput with microfluidic droplet barcoding". In: *Nature biotechnology* 35.7, pp. 640–646. ISSN: 1087-0156. DOI: 10.1038/nbt.3880.

Lan, Freeman, John R. Haliburton, Aaron Yuan, and Adam R. Abate (June 2016). "Droplet barcoding for massively parallel single-molecule deep sequencing". In: *Nature Communications* 7. ISSN: 2041-1723. DOI: 10.1038/ncomms11784.

Lareau, Caleb A. et al. (2019). "Droplet-based combinatorial indexing for massive-scale single-cell chromatin accessibility". eng. In: *Nature Biotechnology* 37.8, pp. 916–924. ISSN: 1546-1696. DOI: 10.1038/s41587-019-0147-6.

Larsen, Simon Asbjørn, Theresa Meldgaard, Agla J. Fridriksdottir, Simon Lykkemark, Pi Camilla Poulsen, Laura F. Overgaard, Helene Bundgaard Petersen, Ole William Petersen, and Peter Kristensen (July 2015). "Selection of a breast cancer subpopulation-specific antibody using phage display on tissue sections". eng. In: *Immunologic Research* 62.3, pp. 263–272. ISSN: 1559-0755. DOI: 10.1007/s12026-015-8657-x.

Le, Xiao-Feng, Amy Lammayot, David Gold, Yiling Lu, Weiqun Mao, Teresa Chang, Adarsh Patel, Gordon B. Mills, and Robert C. Bast (Jan. 2005). "Genes Affecting the Cell Cycle, Growth, Maintenance, and Drug Sensitivity Are Preferentially Regulated by Anti-HER2 Antibody through Phosphatidylinositol 3-Kinase-AKT Signaling". en. In: *Journal of Biological Chemistry* 280.3, pp. 2092–2104. ISSN: 0021-9258, 1083-351X. DOI: 10.1074/jbc.M403080200.

Leader, Benjamin, Quentin J. Baca, and David E. Golan (Jan. 2008). "Protein therapeutics: a summary and pharmacological classification". en. In: *Nature Reviews Drug Discovery* 7.1. Number: 1 Publisher: Nature Publishing Group, pp. 21–39. ISSN: 1474-1784. DOI: 10.1038/nrd2399.

LeBien, Tucker W. and Thomas F. Tedder (Sept. 2008). "B lymphocytes: how they develop and function". en. In: *Blood* 112.5. Publisher: American Society of Hematology, pp. 1570–1580. ISSN: 0006-4971. DOI: 10.1182/blood-2008-02-078071.

Lee, Carol M. Y., Niccolo Iorno, Frederic Sierro, and Daniel Christ (2007). "Selection of human antibody fragments by phage display". eng. In: *Nature Protocols* 2.11, pp. 3001–3008. ISSN: 1750-2799. DOI: 10.1038/nprot.2007.448.

Lee, Rozanne, Mylinh Tran, Mark Nocerini, and Meina Liang (Mar. 2008). "A high-throughput hybridoma selection method using fluorometric microvolume assay technology". eng. In: *Journal of Biomolecular Screening* 13.3, pp. 210–217. ISSN: 1087-0571. DOI: 10.1177/1087057108314148.

Lefrancois, Leo and Douglas S. Lyles (Aug. 1982). "The interaction of antibody with the major surface glycoprotein of vesicular stomatitis virus I. Analysis of neutralizing epitopes with monoclonal antibodies". en. In: *Virology* 121.1, pp. 157–167. ISSN: 0042-6822. DOI: 10.1016/0042-6822(82)90125-8.

Lemal, David M. (Jan. 2004). "Perspective on Fluorocarbon Chemistry". In: *The Journal of Organic Chemistry* 69.1, pp. 1–11. ISSN: 0022-3263. DOI: 10.1021/jo0302556.

Li, Wei Vivian and Jingyi Jessica Li (Mar. 2018). "An accurate and robust imputation method scImpute for single-cell RNA-seq data". en. In: *Nature Communications* 9.1. Number: 1 Publisher: Nature Publishing Group, p. 997. ISSN: 2041-1723. DOI: 10.1038/s41467-018-03405-7.

Li, Ziqiang, Caroline J. Woo, Maria D. Iglesias-Ussel, Diana Ronai, and Matthew D. Scharff (Jan. 2004). "The generation of antibody diversity through somatic hypermutation and class switch recombination". en. In: *Genes & Development* 18.1. Company: Cold Spring Harbor Laboratory Press Distributor: Cold Spring Harbor Laboratory Press Institution: Cold Spring Harbor Laboratory Press Label: Cold Spring Harbor Laboratory Press Publisher: Cold Spring Harbor Lab, pp. 1–11. ISSN: 0890-9369, 1549-5477. DOI: 10.1101/gad.1161904.

Liao, Yang, Gordon K. Smyth, and Wei Shi (Apr. 2014). "featureCounts: an efficient general purpose program for assigning sequence reads to genomic features". eng. In: *Bioinformatics (Oxford, England)* 30.7, pp. 923–930. ISSN: 1367-4811. DOI: 10.1093/bioinformatics/btt656.

Link, D. R., S. L. Anna, D. A. Weitz, and H. A. Stone (Feb. 2004). "Geometrically Mediated Breakup of Drops in Microfluidic Devices". In: *Physical Review Letters* 92.5, p. 054503. DOI: 10.1103/PhysRevLett.92.054503.

Liu, Haolin et al. (Aug. 2015a). "A Rapid Method to Characterize Mouse IgG Antibodies and Isolate Native Antigen Binding IgG B Cell Hybridomas". In: *PLoS ONE* 10.8. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0136613.

Liu, Zongbin, Yeonju Lee, Joon hee Jang, Ying Li, Xin Han, Kenji Yokoi, Mauro Ferrari, Ledu Zhou, and Lidong Qin (Sept. 2015b). "Microfluidic cytometric analysis of cancer cell transportability and invasiveness". eng. In: *Scientific Reports* 5, p. 14272. ISSN: 2045-2322. DOI: 10.1038/srep14272.

Lonberg, Nils (Sept. 2005). "Human antibodies from transgenic animals". eng. In: *Nature Biotechnology* 23.9, pp. 1117–1125. ISSN: 1087-0156. DOI: 10.1038/nbt1135.

Love, J. Christopher, Jehnna L. Ronan, Gijsbert M. Grotenbreg, Annemarthe G. van der Veen, and Hidde L. Ploegh (June 2006). "A microengraving method for rapid selection of single cells producing antigen-specific antibodies". en. In: *Nature Biotechnology* 24.6. Number: 6 Publisher: Nature Publishing Group, pp. 703–707. ISSN: 1546-1696. DOI: 10.1038/nbt1210.

Love, Michael I., Wolfgang Huber, and Simon Anders (Dec. 2014). "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2". In: *Genome Biology* 15.12, p. 550. ISSN: 1474-760X. DOI: 10.1186/s13059-014-0550-8.

Lozzio, C. B. and B. B. Lozzio (Mar. 1975). "Human chronic myelogenous leukemia cell-line with positive Philadelphia chromosome". eng. In: *Blood* 45.3, pp. 321–334. ISSN: 0006-4971.

Lu, Ruei-Min, Yu-Chyi Hwang, I-Ju Liu, Chi-Chiu Lee, Han-Zen Tsai, Hsin-Jung Li, and Han-Chung Wu (Jan. 2020). "Development of therapeutic antibodies for the treatment of diseases". In: *Journal of Biomedical Science* 27.1, p. 1. ISSN: 1423-0127. DOI: 10.1186/s12929-019-0592-z.

Lucchetta, Elena M., Ji Hwan Lee, Lydia A. Fu, Nipam H. Patel, and Rustem F. Ismagilov (Apr. 2005). "Dynamics of Drosophila embryonic patterning network perturbed in space and time using microfluidics". eng. In: *Nature* 434.7037, pp. 1134–1138. ISSN: 1476-4687. DOI: 10.1038/nature03509.

Luecken, Malte D and Fabian J Theis (June 2019). "Current best practices in single-cell RNA-seq analysis: a tutorial". In: *Molecular Systems Biology* 15.6. Publisher: John Wiley & Sons, Ltd, e8746. ISSN: 1744-4292. DOI: 10.15252/msb.20188746.

Lun, Aaron T.L., Davis J. McCarthy, and John C. Marioni (Oct. 2016). "A step-by-step workflow for low-level analysis of single-cell RNA-seq data with Bioconductor". en. In: *F1000Research* 5, p. 2122. ISSN: 2046-1402. DOI: 10.12688/f1000research.9501.2.

Lynch, Mark and Naveen Ramalingam (2019). "Integrated Fluidic Circuits for Single-Cell Omics and Multi-omics Applications". en. In: *Single Molecule and Single Cell Sequencing*. Ed. by Yutaka Suzuki. Advances in Experimental Medicine and Biology. Singapore: Springer, pp. 19–26. ISBN: 9789811360374. DOI: 10.1007/978-981-13-6037-4_2.

Lähnemann, David et al. (Feb. 2020). "Eleven grand challenges in single-cell data science". In: *Genome Biology* 21.1, p. 31. ISSN: 1474-760X. DOI: 10.1186/s13059-020-1926-6.

Maaten, Laurens van der (2014). "Accelerating t-SNE using Tree-Based Algorithms". In: *Journal of Machine Learning Research* 15, pp. 3221–3245.

Maaten, Laurens van der and Geoffrey Hinton (2008). "Visualizing Data using t-SNE". In: *Journal of Machine Learning Research* 9.Nov, pp. 2579–2605. ISSN: ISSN 1533-7928.

Mabonga, Lloyd and Abidemi Paul Kappo (July 2019). "Protein-protein interaction modulators: advances, successes and remaining challenges". In: *Biophysical Reviews* 11.4, pp. 559–581. ISSN: 1867-2450. DOI: 10.1007/s12551-019-00570-x.

Macaulay, Iain C. et al. (June 2015). "G&T-seq: parallel sequencing of single-cell genomes and transcriptomes". en. In: *Nature Methods* 12.6. Number: 6 Publisher: Nature Publishing Group, pp. 519–522. ISSN: 1548-7105. DOI: 10.1038/nmeth.3370.

Macosko, Evan Z. et al. (May 2015). "Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets". In: *Cell* 161.5, pp. 1202–1214. ISSN: 0092-8674. DOI: 10.1016/j.cell.2015.05.002.

MacQueen, J. (1967). "Some methods for classification and analysis of multivariate observations". EN. In: ISSN: 0097-0433. The Regents of the University of California.

Manz, A., Y. Miyahara, J. Miura, Y. Watanabe, H. Miyagi, and K. Sato (Jan. 1990). "Design of an open-tubular column liquid chromatograph using silicon chip technology". en. In: *Sensors and Actuators B: Chemical* 1.1, pp. 249–255. ISSN: 0925-4005. DOI: 10.1016/0925-4005(90)80210-Q.

Mary, Pascaline, Luce Dauphinot, Nadège Bois, Marie-Claude Potier, Vincent Studer, and Patrick Tabeling (June 2011). "Analysis of gene expression at the single-cell level using microdroplet-based microfluidic technology". eng. In: *Biomicrofluidics* 5.2, p. 24109. ISSN: 1932-1058. DOI: 10.1063/1.3596394.

Mathur, Lukas, Martine Ballinger, Ramesh Utharala, and Christoph A. Merten (Nov. 2019). "Microfluidics as an Enabling Technology for Personalized Cancer Therapy". eng. In: *Small (Weinheim an Der Bergstrasse, Germany)*, e1904321. ISSN: 1613-6829. DOI: 10.1002/smll.201904321.

Matsubara, Yasutaka, Kagan Kerman, Masaaki Kobayashi, Shouhei Yamamura, Yasutaka Morita, Yuzuru Takamura, and Eiichi Tamiya (Nov. 2004). "On-chip nanoliter-volume multiplex TaqMan polymerase chain reaction from a single copy based on counting fluorescence released microchambers". eng. In: *Analytical Chemistry* 76.21, pp. 6434–6439. ISSN: 0003-2700. DOI: 10.1021/ac0497149.

Matsue, Tomokazu, Norio Matsumoto, and Isamu Uchida (Jan. 1997). "Rapid micropatterning of living cells by repulsive dielectrophoretic force". en. In: *Electrochimica Acta*. Electrochemical Microsystem Technologies 42.20, pp. 3251–3256. ISSN: 0013-4686. DOI: 10.1016/S0013-4686(97)00175-8.

Mattes, M. Jules (Feb. 2002). "Radionuclide-antibody conjugates for single-cell cytotoxicity". eng. In: *Cancer* 94.4 Suppl, pp. 1215–1223. ISSN: 0008-543X. DOI: 10.1002/cncr.10288.

Matuła, Kinga, Francesca Rivello, and Wilhelm T. S. Huck (2020). "Single-Cell Analysis Using Droplet Microfluidics". en. In: *Advanced Biosystems* 4.1, p. 1900188. ISSN: 2366-7478. DOI: 10.1002/adbi.201900188.

Mazutis, Linas, Jean-Christophe Baret, Patrick Treacy, Yousr Skhiri, Ali Fallah Araghi, Michael Ryckelynck, Valérie Taly, and Andrew D. Griffiths (Oct. 2009). "Multi-step microfluidic droplet processing: kinetic analysis of an in vitro translated enzyme". en. In: *Lab on a Chip* 9.20, pp. 2902–2908. ISSN: 1473-0189. DOI: 10.1039/B907753G.

Mazutis, Linas, John Gilbert, W. Lloyd Ung, David A. Weitz, Andrew D. Griffiths, and John A. Heyman (May 2013). "Single-cell analysis and sorting using droplet-based microfluidics". en. In: *Nature Protocols* 8.5, pp. 870–891. ISSN: 1750-2799. DOI: 10.1038/nprot.2013.046.

McCafferty, John, Andrew D. Griffiths, Greg Winter, and David J. Chiswell (Dec. 1990). "Phage antibodies: filamentous phage displaying antibody variable domains". en. In: *Nature* 348.6301, pp. 552–554. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/348552a0.

McDaniel, Jonathan R., Brandon J. DeKosky, Hidetaka Tanno, Andrew D. Ellington, and George Georgiou (Mar. 2016). "Ultra-high-throughput sequencing of the immune receptor repertoire from millions of lymphocytes". en. In: *Nature Protocols* 11.3. Number: 3 Publisher: Nature Publishing Group, pp. 429–442. ISSN: 1750-2799. DOI: 10.1038/nprot.2016.024.

McInnes, Leland, John Healy, and James Melville (Dec. 2018). "UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction". In: *arXiv:1802.03426 [cs, stat]*. arXiv: 1802.03426.

Mullard, Asher (2019). "2018 FDA drug approvals". eng. In: *Nature Reviews. Drug Discovery* 18.2, pp. 85–89. ISSN: 1474-1784. DOI: 10.1038/d41573-019-00014-x.

Munis, Altar M., Maha Tijani, Mark Hassall, Giada Mattiuzzo, Mary K. Collins, and Yasuhiro Takeuchi (Sept. 2018). "Characterisation of Antibody Interactions with the G Protein of Vesicular Stomatitis Virus Indiana Strain and Other Vesiculovirus G Proteins". en. In: *Journal of Virology*. Publisher: American Society for Microbiology Journals Section: Vaccines and Antiviral Agents. ISSN: 0022-538X, 1098-5514. DOI: 10.1128/JVI.00900-18.

Nemesh, James (Sept. 2018). *Drop-seq Core Computational Protocol V2.0.0*. en.

Niu, Xize, Shelly Gulati, Joshua B. Edel, and Andrew J. deMello (Nov. 2008). "Pillar-induced droplet merging in microfluidic circuits". en. In: *Lab on a Chip* 8.11, pp. 1837–1841. ISSN: 1473-0189. DOI: 10.1039/B813325E.

Niu, Xize, Mengying Zhang, Suili Peng, Weijia Wen, and Ping Sheng (Oct. 2007). "Real-time detection, control, and sorting of microfluidic droplets". In: *Biomicrofluidics* 1.4, p. 044101. DOI: 10.1063/1.2795392.

Norman, Douglas J. (Dec. 1995). "Mechanisms of Action and Overview of OKT3:" en. In: *Therapeutic Drug Monitoring* 17.6, pp. 615–620. ISSN: 0163-4356. DOI: 10.1097/00007691-199512000-00012.

Ogunniyi, Adebola O., Craig M. Story, Eliseo Papa, Eduardo Guillen, and J. Christopher Love (May 2009). "Screening individual hybridomas by microengraving to discover monoclonal antibodies". en. In: *Nature Protocols* 4.5. Number: 5 Publisher: Nature Publishing Group, pp. 767–782. ISSN: 1750-2799. DOI: 10.1038/nprot.2009.40.

Ovacik, Meric and Kedan Lin (Nov. 2018). "Tutorial on Monoclonal Antibody Pharmacokinetics and Its Considerations in Early Development". In: *Clinical and Translational Science* 11.6, pp. 540–552. ISSN: 1752-8054. DOI: 10.1111/cts.12567.

Park, Tai Hyun and Michael L. Shuler (2003). "Integration of Cell Culture and Microfabrication Technology". en. In: *Biotechnology Progress* 19.2. _eprint: https://aiche.onlinelibrary.wiley.com/doi/pdf/10.1021/bp020143k, pp. 243–253. ISSN: 1520-6033. DOI: 10.1021/bp020143k.

Parray, Hilal Ahmed, Shivangi Shukla, Sweety Samal, Tripti Shrivastava, Shubbir Ahmed, Chandresh Sharma, and Rajesh Kumar (Aug. 2020). "Hybridoma technology a versatile method for isolation of monoclonal antibodies, its applicability across species, limitations, advancement and future perspectives". In: *International Immunopharmacology* 85, p. 106639. ISSN: 1567-5769. DOI: 10.1016/j.intimp.2020.106639.

Peled, Jonathan U., Fei Li Kuang, Maria D. Iglesias-Ussel, Sergio Roa, Susan L. Kalis, Myron F. Goodman, and Matthew D. Scharff (2008). "The biochemistry of somatic hypermutation". eng. In: *Annual Review of Immunology* 26, pp. 481–511. ISSN: 0732-0582. DOI: 10.1146/annurev.immunol.26.021607.090236.

Pelfrene, E., M. Mura, A. Cavaleiro Sanches, and M. Cavaleri (Jan. 2019). "Monoclonal antibodies as anti-infective products: a promising future?" en. In: *Clinical Microbiology and Infection* 25.1, pp. 60–64. ISSN: 1198-743X. DOI: 10.1016/j.cmi.2018.04.024.

Pento, Joseph Thomas (Nov. 2017). "Monoclonal Antibodies for the Treatment of Cancer". en. In: *Anticancer Research* 37.11. Publisher: International Institute of Anticancer Research, pp. 5935–5939. ISSN: 0250-7005, 1791-7530.

Peterson, Vanessa M. et al. (Oct. 2017). "Multiplexed quantification of proteins and transcripts in single cells". eng. In: *Nature Biotechnology* 35.10, pp. 936–939. ISSN: 1546-1696. DOI: 10.1038/nbt.3973.

Petukhov, Viktor, Jimin Guo, Ninib Baryawno, Nicolas Severe, David T. Scadden, Maria G. Samsonova, and Peter V. Kharchenko (June 2018). "dropEst: pipeline for accurate estimation of molecular counts in droplet-based single-cell RNA-seq experiments". In: *Genome Biology* 19.1, p. 78. ISSN: 1474-760X. DOI: 10.1186/s13059-018-1449-6.

Pierson, Emma and Christopher Yau (2015). "ZIFA: Dimensionality reduction for zero-inflated single-cell gene expression analysis". In: *Genome Biology* 16. ISSN: 1474-7596. DOI: 10.1186/s13059-015-0805-z.

Piruska, Aigars, Irena Nikcevic, Se Hwan Lee, Chong Ahn, William R. Heineman, Patrick A. Limbach, and Carl J. Seliskar (Dec. 2005). "The autofluorescence of plastic materials and chips measured under laser irradiation". eng. In: *Lab on a Chip* 5.12, pp. 1348–1354. ISSN: 1473-0197. DOI: 10.1039/b508288a.

Poirion, Olivier B., Xun Zhu, Travers Ching, and Lana Garmire (2016). "Single-Cell Transcriptomics Bioinformatics and Computational Challenges". English. In: *Frontiers in Genetics* 7. Publisher: Frontiers. ISSN: 1664-8021. DOI: 10.3389/fgene.2016.00163.

Porter, R. R. (Sept. 1958). "Separation and isolation of fractions of rabbit gamma-globulin containing the antibody and antigenic combining sites". eng. In: *Nature* 182.4636, pp. 670–671. ISSN: 0028-0836. DOI: 10.1038/182670a0.

Protocol - TotalSeq™-A Antibodies and Cell Hashing with 10x Single Cell 3' Reagent Kit v3 3.1 Protocol.

Qin, Dong, Younan Xia, and George M. Whitesides (1996). "Rapid prototyping of complex structures with feature sizes larger than 20 m". In: *Advanced Materials* 8.11, pp. 917–919. ISSN: 1521-4095. DOI: 10.1002/adma.19960081110.

Quake, Stephen R. and Axel Scherer (Nov. 2000). "From Micro- to Nanofabrication with Soft Materials". en. In: *Science* 290.5496, pp. 1536–1540. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.290.5496.1536.

Quan, Phenix-Lan, Martin Sauzade, and Eric Brouzes (Apr. 2018). "dPCR: A Technology Review". In: *Sensors (Basel, Switzerland)* 18.4. ISSN: 1424-8220. DOI: 10.3390/s18041271.

Rajan, Saravanan, Michael R. Kierny, Andrew Mercer, Jincheng Wu, Andrey Tovchigrechko, Herren Wu, William F. DallAcqua, Xiaodong Xiao, and Partha S. Chowdhury (Jan. 2018). "Recombinant human B cell repertoires enable screening for rare, specific, and natively paired antibodies". en. In: *Communications Biology* 1.1. Number: 1 Publisher: Nature Publishing Group, pp. 1–8. ISSN: 2399-3642. DOI: 10.1038/s42003-017-0006-2.

Rayleigh, Lord (1879). "On the Stability, or Instability, of certain Fluid Motions". en. In: *Proceedings of the London Mathematical Society* s1-11.1, pp. 57–72. ISSN: 1460-244X. DOI: 10.1112/plms/s1-11.1.57.

Reddy, Sai T. et al. (Sept. 2010). "Monoclonal antibodies isolated without screening by analyzing the variable-gene repertoire of plasma cells". eng. In: *Nature Biotechnology* 28.9, pp. 965–969. ISSN: 1546-1696. DOI: 10.1038/nbt.1673.

Regev, Aviv et al. (May 2017). "The Human Cell Atlas". en. In: *bioRxiv*. Publisher: Cold Spring Harbor Laboratory Section: New Results, p. 121202. DOI: 10.1101/121202.

Replogle, Joseph M. et al. (Mar. 2020). "Combinatorial single-cell CRISPR screens by direct guide RNA capture and targeted sequencing". en. In: *Nature Biotechnology*. Publisher: Nature Publishing Group, pp. 1–8. ISSN: 1546-1696. DOI: 10.1038/s41587-020-0470-y.

Risso, Davide, John Ngai, Terence P. Speed, and Sandrine Dudoit (Sept. 2014). "Normalization of RNA-seq data using factor analysis of control genes or samples". en. In: *Nature Biotechnology* 32.9. Number: 9 Publisher: Nature Publishing Group, pp. 896–902. ISSN: 1546-1696. DOI: 10.1038/nbt.2931.

Risso, Davide, Fanny Perraudeau, Svetlana Gribkova, Sandrine Dudoit, and Jean-Philippe Vert (Jan. 2018). "A general and flexible method for signal extraction from single-cell RNA-seq data". en. In: *Nature Communications* 9.1. Number: 1 Publisher: Nature Publishing Group, p. 284. ISSN: 2041-1723. DOI: 10.1038/s41467-017-02554-5.

Roelli, Patrick (Sept. 2020). *Hoohm/dropSeqPipe*. original-date: 2016-08-02T10:03:44Z.

Rognes, Torbjørn, Tomáš Flouri, Ben Nichols, Christopher Quince, and Frédéric Mahé (2016). "VSEARCH: a versatile open source tool for metagenomics". eng. In: *PeerJ* 4, e2584. ISSN: 2167-8359. DOI: 10.7717/peerj.2584.

Rosenfeld, Nitzan, Jonathan W. Young, Uri Alon, Peter S. Swain, and Michael B. Elowitz (Mar. 2005). "Gene Regulation at the Single-Cell Level". en. In: *Science* 307.5717, pp. 1962–1965. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.1106914.

Rotem, Assaf, Oren Ram, Noam Shoresh, Ralph A. Sperling, Alon Goren, David A. Weitz, and Bradley E. Bernstein (Nov. 2015). "Single-cell ChIP-seq reveals cell subpopulations defined by chromatin state". eng. In: *Nature Biotechnology* 33.11, pp. 1165–1172. ISSN: 1546-1696. DOI: 10.1038/nbt.3383.

Ruppen, Janine, Franziska D. Wildhaber, Christoph Strub, Sean R. R. Hall, Ralph A. Schmid, Thomas Geiser, and Olivier T. Guenat (June 2015). "Towards personalized medicine: chemosensitivity assays of patient lung cancer cell spheroids in a perfused microfluidic platform". en. In: *Lab on a Chip* 15.14, pp. 3076–3085. ISSN: 1473-0189. DOI: 10.1039/C5LC00454C.

Ryckelynck, Michael (May 2020). *thesis*.

Ryman, Josiah T. and Bernd Meibohm (Sept. 2017). "Pharmacokinetics of Monoclonal Antibodies: Pharmacokinetics of Monoclonal Antibodies". en. In: *CPT: Pharmacometrics & Systems Pharmacology* 6.9, pp. 576–588. ISSN: 21638306. DOI: 10.1002/psp4.12224.

Saadatpour, Assieh, Shujing Lai, Guoji Guo, and Guo-Cheng Yuan (Oct. 2015). "Single-cell analysis in cancer genomics". In: *Trends in genetics : TIG* 31.10, pp. 576–586. ISSN: 0168-9525. DOI: 10.1016/j.tig.2015.07.003.

Sackmann, Eric K., Anna L. Fulton, and David J. Beebe (Mar. 2014). "The present and future role of microfluidics in biomedical research". en. In: *Nature* 507.7491, pp. 181–189. ISSN: 1476-4687. DOI: 10.1038/nature13118.

Saiki, R. K., S. Scharf, F. Faloona, K. B. Mullis, G. T. Horn, H. A. Erlich, and N. Arnheim (Dec. 1985). "Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia". en. In: *Science* 230.4732. Publisher: American Association for the Advancement of Science Section: Articles, pp. 1350–1354. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.2999980.

Saikia, Mridusmita et al. (Jan. 2019). "Simultaneous multiplexed amplicon sequencing and transcriptome profiling in single cells". en. In: *Nature Methods* 16.1. Number: 1 Publisher: Nature Publishing Group, pp. 59–62. ISSN: 1548-7105. DOI: 10.1038/s41592-018-0259-9.

Salazar, Georgina, Ningyan Zhang, Tong-Ming Fu, and Zhiqiang An (July 2017). "Antibody therapies for the prevention and treatment of viral infections". en. In: *npj Vaccines* 2.1. Number: 1 Publisher: Nature Publishing Group, pp. 1–12. ISSN: 2059-0105. DOI: 10.1038/s41541-017-0019-3.

Sarkar, Saheli, Vinny Motwani, Pooja Sabhachandani, Noa Cohen, and Tania Konry (June 2015). "T Cell Dynamic Activation and Functional Analysis in Nanoliter Droplet Microarray". eng. In: *Journal of Clinical & Cellular Immunology* 6.3. DOI: 10.4172/2155-9899.1000334.

Satija, Rahul, Jeffrey A. Farrell, David Gennert, Alexander F. Schier, and Aviv Regev (May 2015). "Spatial reconstruction of single-cell gene expression data". en. In: *Nature Biotechnology* 33.5. Number: 5 Publisher: Nature Publishing Group, pp. 495–502. ISSN: 1546-1696. DOI: 10.1038/nbt.3192.

Sauzade, M. and E. Brouzes (2017). "Deterministic trapping, encapsulation and retrieval of single-cells". en. In: *Lab on a Chip* 17.13, pp. 2186–2192. ISSN: 1473-0197, 1473-0189. DOI: 10.1039/C7LC00283A.

Scheid, Johannes F. et al. (Apr. 2009). "A method for identification of HIV gp140 binding memory B cells in human blood". eng. In: *Journal of Immunological Methods* 343.2, pp. 65–67. ISSN: 1872-7905. DOI: 10.1016/j.jim.2008.11.012.

Schneider, Ulrich, Hans-Ulrich Schwenk, and Georg Bornkamm (1977). "Characterization of EBV-genome negative "null" and "T" cell lines derived from children with acute lymphoblastic leukemia and leukemic transformed non-Hodgkin lymphoma". en. In: *International Journal of Cancer* 19.5. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/ijc.2910190505, pp. 621–626. ISSN: 1097-0215. DOI: 10.1002/ijc.2910190505.

Schoeman, Rogier M., Evelien W. M. Kemna, Floor Wolbers, and Albert van den Berg (2014). "High-throughput deterministic single-cell encapsulation and droplet pairing, fusion, and shrinkage in a single microfluidic device". en. In: *ELECTROPHORESIS*

35.2-3. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/elps.201300179, pp. 385–392. ISSN: 1522-2683. DOI: 10.1002/elps.201300179.

Schraivogel, Daniel et al. (June 2020). "Targeted Perturb-seq enables genome-scale genetic screens in single cells". en. In: *Nature Methods* 17.6, pp. 629–635. ISSN: 1548-7091, 1548-7105. DOI: 10.1038/s41592-020-0837-5.

Schroeder, Harry W Jr. and Lisa Cavacini (2010). "Structure and Function of Immunoglobulins". In:

Scott, Andrew M., James P. Allison, and Jedd D. Wolchok (May 2012). "Monoclonal antibodies in cancer therapy". In: *Cancer Immunity* 12. ISSN: 1424-9634.

Seah, Yu Fen Samantha, Hongxing Hu, and Christoph A. Merten (Feb. 2018). "Microfluidic single-cell technology in immunology and antibody screening". en. In: *Molecular Aspects of Medicine*. The emerging field of single-cell analysis 59, pp. 47–61. ISSN: 0098-2997. DOI: 10.1016/j.mam.2017.09.004.

Segaliny, Aude I., Guideng Li, Lingshun Kong, Ci Ren, Xiaoming Chen, Jessica K. Wang, David Baltimore, Guikai Wu, and Weian Zhao (Dec. 2018). "Functional TCR T cell screening using single-cell droplet microfluidics". en. In: *Lab on a Chip* 18.24. Publisher: The Royal Society of Chemistry, pp. 3733–3749. ISSN: 1473-0189. DOI: 10.1039/C8LC00818C.

Shahi, Payam, Samuel C. Kim, John R. Haliburton, Zev J. Gartner, and Adam R. Abate (Mar. 2017). "Abseq: Ultrahigh-throughput single cell protein profiling with droplet microfluidic barcoding". en. In: *Scientific Reports* 7.1. Number: 1 Publisher: Nature Publishing Group, p. 44447. ISSN: 2045-2322. DOI: 10.1038/srep44447.

Shao, Keke, Weifeng Ding, Feng Wang, Haiquan Li, Da Ma, and Huimin Wang (Sept. 2011). "Emulsion PCR: A High Efficient Way of PCR Amplification of Random DNA Libraries in Aptamer Selection". en. In: *PLOS ONE* 6.9. Publisher: Public Library of Science, e24910. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0024910.

Sheehan, Jared and Wayne A. Marasco (Feb. 2015). "Phage and Yeast Display". en. In: *Microbiology Spectrum* 3.1. Company: asm Pub2Web Distributor: asm Pub2Web Institution: asm Pub2Web Label: asm Pub2Web Publisher: American Society of Microbiology. ISSN: 2165-0497. DOI: 10.1128/microbiolspec.AID-0028-2014.

Shembekar, Nachiket, Chawaree Chaipan, Ramesh Utharala, and Christoph A. Merten (Apr. 2016). "Droplet-based microfluidics in drug discovery, transcriptomics and high-throughput molecular genetics". en. In: *Lab on a Chip* 16.8, pp. 1314–1331. ISSN: 1473-0189. DOI: 10.1039/C6LC00249H.

Shembekar, Nachiket, Hongxing Hu, David Eustace, and Christoph A. Merten (Feb. 2018). "Single-Cell Droplet Microfluidic Screening for Antibodies Specifically Binding to Target Cells". en. In: *Cell Reports* 22.8, pp. 2206–2215. ISSN: 2211-1247. DOI: 10.1016/j.celrep.2018.01.071.

Shum, Eleen Y., Elisabeth M. Walczak, Christina Chang, and H. Christina Fan (2019). "Quantitation of mRNA Transcripts and Proteins Using the BD Rhapsody™ Single-Cell Analysis System". en. In: *Single Molecule and Single Cell Sequencing*. Ed. by Yutaka Suzuki. Advances in Experimental Medicine and Biology. Singapore: Springer, pp. 63–79. ISBN: 9789811360374. DOI: 10.1007/978-981-13-6037-4_5.

Sin, Aaron, Katherine C. Chin, Muhammad F. Jamil, Yordan Kostov, Govind Rao, and Michael L. Shuler (2004). "The Design and Fabrication of Three-Chamber Microscale Cell Culture Analog Devices with Integrated Dissolved Oxygen Sensors". en. In: *Biotechnology Progress* 20.1. _eprint: https://aiche.onlinelibrary.wiley.com/doi/pdf/10.1021/bp034077d, pp. 338–345. ISSN: 1520-6033. DOI: 10.1021/bp034077d.

Singh, Mandeep et al. (July 2019). "High-throughput targeted long-read single cell sequencing reveals the clonal and transcriptional landscape of lymphocytes". en. In: *Nature Communications* 10.1. Number: 1 Publisher: Nature Publishing Group, p. 3120. ISSN: 2041-1723. DOI: 10.1038/s41467-019-11049-4.

Singhal, Anupam, Charles A. Haynes, and Carl L. Hansen (Oct. 2010). "Microfluidic measurement of antibody-antigen binding kinetics from low-abundance samples and

single cells". eng. In: *Analytical Chemistry* 82.20, pp. 8671–8679. ISSN: 1520-6882. DOI: 10.1021/ac101956e.

Single Cell Immune Profiling. Library Catalog: www.10xgenomics.com.

Skelley, Alison M., Oktay Kirak, Heikyung Suh, Rudolf Jaenisch, and Joel Voldman (Feb. 2009). "Microfluidic control of cell pairing and fusion". en. In: *Nature Methods* 6.2, pp. 147–152. ISSN: 1548-7105. DOI: 10.1038/nmeth.1290.

Slastnikova, Tatiana A., A. V. Ulasov, A. A. Rosenkranz, and A. S. Sobolev (Oct. 2018). "Targeted Intracellular Delivery of Antibodies: The State of the Art". In: *Frontiers in Pharmacology* 9. ISSN: 1663-9812. DOI: 10.3389/fphar.2018.01208.

Smeets, Ruben L. et al. (Mar. 2012). "Molecular pathway profiling of T lymphocyte signal transduction pathways; Th1 and Th2 genomic fingerprints are defined by TCR and CD28-mediated signaling". In: *BMC Immunology* 13.1, p. 12. ISSN: 1471-2172. DOI: 10.1186/1471-2172-13-12.

Smith, G. P. (June 1985). "Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface". eng. In: *Science (New York, N.Y.)* 228.4705, pp. 1315–1317. ISSN: 0036-8075. DOI: 10.1126/science.4001944.

Soneson, Charlotte and Mark D. Robinson (Apr. 2018). "Bias, robustness and scalability in single-cell differential expression analysis". en. In: *Nature Methods* 15.4. Number: 4 Publisher: Nature Publishing Group, pp. 255–261. ISSN: 1548-7105. DOI: 10.1038/nmeth.4612.

Soule, H. D., J. Vazquez, A. Long, S. Albert, and M. Brennan (Nov. 1973). "A Human Cell Line From a Pleural Effusion Derived From a Breast Carcinoma". en. In: *JNCI: Journal of the National Cancer Institute* 51.5. Publisher: Oxford Academic, pp. 1409–1416. ISSN: 0027-8874. DOI: 10.1093/jnci/51.5.1409.

Sousa, Isabel Garcia et al. (July 2019). "Gene expression profile of human T cells following a single stimulation of peripheral blood mononuclear cells with anti-CD3 antibodies". In: *BMC Genomics* 20. ISSN: 1471-2164. DOI: 10.1186/s12864-019-5967-8.

Spencer, Sarah J. et al. (Feb. 2016). "Massively parallel sequencing of single cells by epicPCR links functional genes with phylogenetic markers". en. In: *The ISME Journal* 10.2. Number: 2 Publisher: Nature Publishing Group, pp. 427–436. ISSN: 1751-7370. DOI: 10.1038/ismej.2015.124.

Squires, Todd M. and Stephen R. Quake (Oct. 2005). "Microfluidics: Fluid physics at the nanoliter scale". en. In: *Reviews of Modern Physics* 77.3, pp. 977–1026. ISSN: 0034-6861, 1539-0756. DOI: 10.1103/RevModPhys.77.977.

Stegle, Oliver, Sarah A. Teichmann, and John C. Marioni (Mar. 2015). "Computational and analytical challenges in single-cell transcriptomics". en. In: *Nature Reviews Genetics* 16.3. Number: 3 Publisher: Nature Publishing Group, pp. 133–145. ISSN: 1471-0064. DOI: 10.1038/nrg3833.

Stoeckius, Marlon, Christoph Hafemeister, William Stephenson, Brian Houck-Loomis, Pratip K. Chattopadhyay, Harold Swerdlow, Rahul Satija, and Peter Smibert (Mar. 2017a). *Large-scale simultaneous measurement of epitopes and transcriptomes in single cells*. en. preprint. Genomics. DOI: 10.1101/113068.

— (Sept. 2017b). "Simultaneous epitope and transcriptome measurement in single cells". en. In: *Nature Methods* 14.9, pp. 865–868. ISSN: 1548-7105. DOI: 10.1038/nmeth.4380.

Stoeckius, Marlon, Shiwei Zheng, Brian Houck-Loomis, Stephanie Hao, Bertrand Z. Yeung, William M. Mauck, Peter Smibert, and Rahul Satija (Dec. 2018). "Cell Hashing with barcoded antibodies enables multiplexing and doublet detection for single cell genomics". In: *Genome Biology* 19.1, p. 224. ISSN: 1474-760X. DOI: 10.1186/s13059-018-1603-1.

Story, Craig M., Eliseo Papa, Chih-Chi Andrew Hu, Jehnna L. Ronan, Kara Herlihy, Hidde L. Ploegh, and J. Christopher Love (Nov. 2008). "Profiling antibody responses by multiparametric analysis of primary B cells". en. In: *Proceedings of the National Academy of Sciences* 105.46. Publisher: National Academy of Sciences Section: Physical Sciences, pp. 17902–17907. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.0805470105.

Streets, Aaron M. et al. (May 2014). "Microfluidic single-cell whole-transcriptome sequencing". In: *Proceedings of the National Academy of Sciences of the United States of America* 111.19, pp. 7048–7053. ISSN: 0027-8424. DOI: 10.1073/pnas.1402030111.

"5 - Sources of antibody variable chains" (Jan. 2012). en. In: *Therapeutic Antibody Engineering*. Ed. by William R. Strohl and Lila M. Strohl. Woodhead Publishing Series in Biomedicine. Woodhead Publishing, pp. 77–595. ISBN: 978-1-907568-37-4. DOI: 10.1533/9781908818096.77.

Subramanian, Aravind et al. (Nov. 2017). "A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles". eng. In: *Cell* 171.6, 1437–1452.e17. ISSN: 1097-4172. DOI: 10.1016/j.cell.2017.10.049.

Sudo, Ryo, Seok Chung, Ioannis K. Zervantonakis, Vernella Vickerman, Yasuko Toshimitsu, Linda G. Griffith, and Roger D. Kamm (July 2009). "Transport-mediated angiogenesis in 3D epithelial coculture". eng. In: *FASEB journal: official publication of the Federation of American Societies for Experimental Biology* 23.7, pp. 2155–2164. ISSN: 1530-6860. DOI: 10.1096/fj.08-122820.

Sumida, Takeshi, Hiroshi Yanagawa, and Nobuhide Doi (Sept. 2012). *In Vitro Selection of Fab Fragments by mRNA Display and Gene-Linking Emulsion PCR*. en. Research Article. ISSN: 2090-0201 Library Catalog: www.hindawi.com Pages: e371379 Publisher: Hindawi Volume: 2012. DOI: https://doi.org/10.1155/2012/371379.

Sutherland, R., D. Delia, C. Schneider, R. Newman, J. Kemshead, and M. Greaves (July 1981). "Ubiquitous cell-surface glycoprotein on tumor cells is proliferation-associated receptor for transferrin". en. In: *Proceedings of the National Academy of Sciences* 78.7. Publisher: National Academy of Sciences Section: Research Article, pp. 4515–4519. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.78.7.4515.

Svec, David, Ales Tichopad, Vendula Novosadova, Michael W. Pfaffl, and Mikael Kubista (Mar. 2015). "How good is a PCR efficiency estimate: Recommendations for precise and robust qPCR efficiency assessments". eng. In: *Biomolecular Detection and Quantification* 3, pp. 9–16. ISSN: 2214-7535. DOI: 10.1016/j.bdq.2015.01.005.

Svensson, Valentine, Eduardo da Veiga Beltrame, and Lior Pachter (Sept. 2019). *Quantifying the tradeoff between sequencing depth and cell number in single-cell RNA-seq*. en. preprint. Genomics. DOI: 10.1101/762773.

Sykes, P. J., S. H. Neoh, M. J. Brisco, E. Hughes, J. Condon, and A. A. Morley (Sept. 1992). "Quantitation of targets for PCR by use of limiting dilution". eng. In: *BioTechniques* 13.3, pp. 444–449. ISSN: 0736-6205.

Tambe, Akshay and Lior Pachter (Jan. 2019). "Barcode identification for single cell genomics". In: *BMC Bioinformatics* 20.1, p. 32. ISSN: 1471-2105. DOI: 10.1186/s12859-019-2612-0.

Tatosian, Daniel A. and Michael L. Shuler (2009). "A novel system for evaluation of drug mixtures for potential efficacy in treating multidrug resistant cancers". en. In: *Biotechnology and Bioengineering* 103.1. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/bit.22219, pp. 187–198. ISSN: 1097-0290. DOI: 10.1002/bit.22219.

Taylor, Sean C., Genevieve Laperriere, and Hugo Germain (May 2017). "Droplet Digital PCR versus qPCR for gene expression analysis with low abundant targets: from variable nonsense to publication quality data". en. In: *Scientific Reports* 7.1. Number: 1 Publisher: Nature Publishing Group, p. 2409. ISSN: 2045-2322. DOI: 10.1038/s41598-017-02217-x.

Thorens, B., A. Porret, L. Bühler, S. P. Deng, P. Morel, and C. Widmann (Nov. 1993). "Cloning and functional expression of the human islet GLP-1 receptor. Demonstration that exendin-4 is an agonist and exendin-(9-39) an antagonist of the receptor". eng. In: *Diabetes* 42.11, pp. 1678–1682. ISSN: 0012-1797. DOI: 10.2337/diab.42.11.1678.

Thorsen, T., R. W. Roberts, F. H. Arnold, and S. R. Quake (Apr. 2001). "Dynamic pattern formation in a vesicle-generating microfluidic device". eng. In: *Physical Review Letters* 86.18, pp. 4163–4166. ISSN: 0031-9007. DOI: 10.1103/PhysRevLett.86.4163.

Tickle, Simon, Ralph Adams, Derek Brown, Meryn Griffiths, Daniel Lightwood, and Alastair Lawson (Oct. 2009). "High-Throughput Screening for High Affinity Antibodies". en.

In: *Journal of the Association for Laboratory Automation* 14.5, pp. 303–307. ISSN: 15355535. DOI: 10.1016/j.jala.2009.05.004.

Tiller, Thomas, Eric Meffre, Sergey Yurasov, Makoto Tsuiji, Michel C. Nussenzweig, and Hedda Wardemann (Jan. 2008). "Efficient generation of monoclonal antibodies from single human B cells by single cell RT-PCR and expression vector cloning". In: *Journal of immunological methods* 329.1-2, pp. 112–124. ISSN: 0022-1759. DOI: 10.1016/j.jim.2007.09.017.

Tonegawa, Susumu (Apr. 1983). "Somatic generation of antibody diversity". en. In: *Nature* 302.5909, pp. 575–581. ISSN: 1476-4687. DOI: 10.1038/302575a0.

Traggiai, Elisabetta, Stephan Becker, Kanta Subbarao, Larissa Kolesnikova, Yasushi Uematsu, Maria Rita Gismondo, Brian R. Murphy, Rino Rappuoli, and Antonio Lanzavecchia (Aug. 2004). "An efficient method to make human monoclonal antibodies from memory B cells: potent neutralization of SARS coronavirus". eng. In: *Nature Medicine* 10.8, pp. 871–875. ISSN: 1078-8956. DOI: 10.1038/nm1080.

Trenevska, Iva, Demin Li, and Alison H. Banham (Aug. 2017). "Therapeutic Antibodies against Intracellular Tumor Antigens". In: *Frontiers in Immunology* 8. ISSN: 1664-3224. DOI: 10.3389/fimmu.2017.01001.

Umbanhowar, P. B., V. Prasad, and D. A. Weitz (Jan. 2000). "Monodisperse Emulsion Generation via Drop Break Off in a Coflowing Stream". en. In: *Langmuir* 16.2, pp. 347–351. ISSN: 0743-7463, 1520-5827. DOI: 10.1021/la990101e.

Unger, M. A., H. P. Chou, T. Thorsen, A. Scherer, and S. R. Quake (Apr. 2000). "Monolithic microfabricated valves and pumps by multilayer soft lithography". eng. In: *Science (New York, N.Y.)* 288.5463, pp. 113–116. ISSN: 0036-8075. DOI: 10.1126/science.288.5463.113.

Urquhart, Lisa (Mar. 2020a). "Top companies and drugs by sales in 2019". en. In: *Nature Reviews Drug Discovery* 19.4. Number: 4 Publisher: Nature Publishing Group, pp. 228–228. DOI: 10.1038/d41573-020-00047-7.

— (Jan. 2020b). "Top product forecasts for 2020". en. In: *Nature Reviews Drug Discovery* 19.2. Number: 2 Publisher: Nature Publishing Group, pp. 86–86. DOI: 10.1038/d41573-020-00011-5.

Vallejo, Andres F., James Davies, Amit Grover, Ching-Hsuan Tsai, Robert Jepras, Marta E. Polak, and Jonathan West (Oct. 2019). *Resolving cellular systems by ultra-sensitive and economical single-cell transcriptome filtering*. en. preprint. Molecular Biology. DOI: 10.1101/800631.

Vaughan, T. J. et al. (Mar. 1996). "Human antibodies with sub-nanomolar affinities isolated from a large non-immunized phage display library". eng. In: *Nature Biotechnology* 14.3, pp. 309–314. ISSN: 1087-0156. DOI: 10.1038/nbt0396-309.

Velve-Casquillas, Guilhem, Maël Le Berre, Matthieu Piel, and Phong T. Tran (Feb. 2010). "Microfluidic tools for cell biological research". In: *Nano today* 5.1, pp. 28–47. ISSN: 1748-0132. DOI: 10.1016/j.nantod.2009.12.001.

Vieth, Beate, Swati Parekh, Christoph Ziegenhain, Wolfgang Enard, and Ines Hellmann (Oct. 2019). "A systematic evaluation of single cell RNA-seq analysis pipelines". en. In: *Nature Communications* 10.1. Number: 1 Publisher: Nature Publishing Group, p. 4667. ISSN: 2041-1723. DOI: 10.1038/s41467-019-12266-7.

Villani, Alexandra-Chloé et al. (2017). "Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors". eng. In: *Science (New York, N.Y.)* 356.6335. ISSN: 1095-9203. DOI: 10.1126/science.aah4573.

Vogelstein, Bert and Kenneth W. Kinzler (Aug. 1999). "Digital PCR". en. In: *Proceedings of the National Academy of Sciences* 96.16. Publisher: National Academy of Sciences Section: Biological Sciences, pp. 9236–9241. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.96.16.9236.

Vu, Timothy Quang, Ricardo Miguel Bessa de Castro, and Lidong Qin (Mar. 2017). "Bridging the gap: Microfluidic devices for short and long distance cell-cell communication". In: *Lab on a chip* 17.6, pp. 1009–1023. ISSN: 1473-0197. DOI: 10.1039/c6lc01367h.

Wagner, Olaf, Julian Thiele, Marie Weinhart, Linas Mazutis, David A. Weitz, Wilhelm T. S. Huck, and Rainer Haag (Dec. 2015). "Biocompatible fluorinated polyglycerols for droplet microfluidics as an alternative to PEG-based copolymer surfactants". en. In: *Lab on a Chip* 16.1, pp. 65–69. ISSN: 1473-0189. DOI: 10.1039/C5LC00823A.

Wan, Hong (Mar. 2016). "An Overall Comparison of Small Molecules and Large Biologics in ADME Testing". en. In: *ADMET and DMPK* 4.1. Number: 1, pp. 1–22. ISSN: 1848-7718. DOI: 10.5599/admet.4.1.276.

Wang, Jianbin, H. Christina Fan, Barry Behr, and Stephen R. Quake (July 2012). "Genome-Wide Single-Cell Analysis of Recombination Activity and de novo Mutation Rates in Human Sperm". In: *Cell* 150.2, pp. 402–412. ISSN: 0092-8674. DOI: 10.1016/j.cell.2012.06.030.

Wang, Tianyu, Boyang Li, Craig E. Nelson, and Sheida Nabavi (Jan. 2019). "Comparative analysis of differential gene expression analysis tools for single-cell RNA sequencing data". In: *BMC Bioinformatics* 20.1, p. 40. ISSN: 1471-2105. DOI: 10.1186/s12859-019-2599-6.

Wang, W., E. Q. Wang, and J. P. Balthasar (2008). "Monoclonal Antibody Pharmacokinetics and Pharmacodynamics". en. In: *Clinical Pharmacology & Therapeutics* 84.5. _eprint: https://ascpt.onlinelibrary.wiley.com/doi/pdf/10.1038/clpt.2008.170, pp. 548–558. ISSN: 1532-6535. DOI: 10.1038/clpt.2008.170.

Weinberger, Leor S., John C. Burnett, Jared E. Toettcher, Adam P. Arkin, and David V. Schaffer (July 2005). "Stochastic gene expression in a lentiviral positive-feedback loop: HIV-1 Tat fluctuations drive phenotypic diversity". eng. In: *Cell* 122.2, pp. 169–182. ISSN: 0092-8674. DOI: 10.1016/j.cell.2005.06.006.

Weiner, Louis M., Rishi Surana, and Shangzi Wang (May 2010). "Antibodies and cancer therapy: versatile platforms for cancer immunotherapy". In: *Nature reviews. Immunology* 10.5, pp. 317–327. ISSN: 1474-1733. DOI: 10.1038/nri2744.

Weissman, A. M., R. D. Klausner, K. Rao, and J. B. Harford (Mar. 1986). "Exposure of K562 cells to anti-receptor monoclonal antibody OKT9 results in rapid redistribution and enhanced degradation of the transferrin receptor". eng. In: *The Journal of Cell Biology* 102.3, pp. 951–958. ISSN: 0021-9525.

Welch, Danny R. et al. (1991). "Characterization of a highly invasive and spontaneously metastatic human malignant melanoma cell line". en. In: *International Journal of Cancer* 47.2. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/ijc.2910470211, pp. 227–237. ISSN: 1097-0215. DOI: 10.1002/ijc.2910470211.

Wetterstrand, K. A. *The Cost of Sequencing a Human Genome*. en.

Whale, Alexandra S., Jim F. Huggett, and Svilen Tzonev (Dec. 2016). "Fundamentals of multiplexing with digital PCR". en. In: *Biomolecular Detection and Quantification*. Special issue on dPCR 10, pp. 15–23. ISSN: 2214-7535. DOI: 10.1016/j.bdq.2016.05.002.

Whitesides, George M. (July 2006). "The origins and the future of microfluidics". en. In: *Nature* 442.7101, pp. 368–373. ISSN: 1476-4687. DOI: 10.1038/nature05058.

Williams, Richard, Sergio G. Peisajovich, Oliver J. Miller, Shlomo Magdassi, Dan S. Tawfik, and Andrew D. Griffiths (July 2006). "Amplification of complex gene libraries by emulsion PCR". en. In: *Nature Methods* 3.7. Number: 7 Publisher: Nature Publishing Group, pp. 545–550. ISSN: 1548-7105. DOI: 10.1038/nmeth896.

Wippold, Jose A., Han Wang, Joseph Tingling, Julian L. Leibowitz, Paul de Figueiredo, and Arum Han (May 2020). "PRESCIENT: platform for the rapid evaluation of antibody success using integrated microfluidics enabled technology". en. In: *Lab on a Chip* 20.9. Publisher: The Royal Society of Chemistry, pp. 1628–1638. ISSN: 1473-0189. DOI: 10.1039/C9LC01165J.

Wolf, F. Alexander, Philipp Angerer, and Fabian J. Theis (Feb. 2018). "SCANPY: large-scale single-cell gene expression data analysis". In: *Genome Biology* 19.1, p. 15. ISSN: 1474-760X. DOI: 10.1186/s13059-017-1382-0.

Wright, S. D., R. A. Ramos, P. S. Tobias, R. J. Ulevitch, and J. C. Mathison (Sept. 1990). "CD14, a receptor for complexes of lipopolysaccharide (LPS) and LPS binding protein". eng.

In: *Science (New York, N.Y.)* 249.4975, pp. 1431–1433. ISSN: 0036-8075. DOI: 10.1126/science.1698311.

Xiao, Xiaodong et al. (Aug. 2017). "A high-throughput platform for population reformatting and mammalian expression of phage display libraries to enable functional screening as full-length IgG". en. In: *mAbs* 9.6, pp. 996–1006. ISSN: 1942-0862, 1942-0870. DOI: 10.1080/19420862.2017.1337617.

Xie, Shiqi, Jialei Duan, Boxun Li, Pei Zhou, and Gary C. Hon (Apr. 2017). "Multiplexed Engineering and Analysis of Combinatorial Enhancer Activity in Single Cells". en. In: *Molecular Cell* 66.2, 285–299.e5. ISSN: 1097-2765. DOI: 10.1016/j.molcel.2017.03.007.

Xu, Tao, Wanqing Yue, Cheuk-Wing Li, Xinsheng Yao, Guoping Cai, and Mengsu Yang (Sept. 2010). "Real-time monitoring of suspension cell-cell communication using an integrated microfluidics". eng. In: *Lab on a Chip* 10.17, pp. 2271–2278. ISSN: 1473-0197. DOI: 10.1039/c004844e.

Yamanaka, Yvonne J., Christoph T. Berger, Magdalena Sips, Patrick C. Cheney, Galit Alter, and J. Christopher Love (Sept. 2012). "Single-cell analysis of the dynamics and functional outcomes of interactions between human natural killer cells and target cells". en. In: *Integrative Biology* 4.10. Publisher: Oxford Academic, pp. 1175–1184. DOI: 10.1039/c2ib20167d.

Yamashita, Makiko, Yoshinroi Katakura, and Sanetaka Shirahata (Dec. 2007). "Recent advances in the generation of human monoclonal antibody". In: *Cytotechnology* 55.2-3, pp. 55–60. ISSN: 0920-9069. DOI: 10.1007/s10616-007-9072-5.

Ye, Jian, Ning Ma, Thomas L. Madden, and James M. Ostell (July 2013). "IgBLAST: an immunoglobulin variable domain sequence analysis tool". In: *Nucleic Acids Research* 41.Web Server issue, W34–W40. ISSN: 0305-1048. DOI: 10.1093/nar/gkt382.

Young, Matthew D., Matthew J. Wakefield, Gordon K. Smyth, and Alicia Oshlack (Feb. 2010). "Gene ontology analysis for RNA-seq: accounting for selection bias". In: *Genome Biology* 11.2, R14. ISSN: 1474-760X. DOI: 10.1186/gb-2010-11-2-r14.

Zahavi, Eitan Erez, Ariel Ionescu, Shani Gluska, Tal Gradus, Keren Ben-Yaakov, and Eran Perlson (Mar. 2015). "A compartmentalized microfluidic neuromuscular co-culture system reveals spatial aspects of GDNF functions". eng. In: *Journal of Cell Science* 128.6, pp. 1241–1252. ISSN: 1477-9137. DOI: 10.1242/jcs.167544.

Zanoni, Ivan, Renato Ostuni, Lorri R. Marek, Simona Barresi, Roman Barbalat, Gregory M. Barton, Fancesca Granucci, and Jonathan C. Kagan (Nov. 2011). "CD14 controls the LPS-induced endocytosis of Toll-like Receptor 4". In: *Cell* 147.4, pp. 868–880. ISSN: 0092-8674. DOI: 10.1016/j.cell.2011.09.051.

Zeisel, Amit et al. (2018). "Molecular Architecture of the Mouse Nervous System". eng. In: *Cell* 174.4, 999–1014.e22. ISSN: 1097-4172. DOI: 10.1016/j.cell.2018.06.021.

Zervantonakis, Ioannis K., Chandrasekhar R. Kothapalli, Seok Chung, Ryo Sudo, and Roger D. Kamm (Mar. 2011). "Microfluidic devices for studying heterotypic cell-cell interactions and tissue specimen cultures under controlled microenvironments". In: *Biomicrofluidics* 5.1. ISSN: 1932-1058. DOI: 10.1063/1.3553237.

Zhang, Hongkai et al. (Dec. 2015). "Autocrine selection of a GLP-1R G-protein biased agonist with potent antidiabetic effects". eng. In: *Nature Communications* 6, p. 8918. ISSN: 2041-1723. DOI: 10.1038/ncomms9918.

Zhang, Huifa, Gareth Jenkins, Yuan Zou, Zhi Zhu, and Chaoyong James Yang (Apr. 2012). "Massively parallel single-molecule and single-cell emulsion reverse transcription polymerase chain reaction using agarose droplet microfluidics". eng. In: *Analytical Chemistry* 84.8, pp. 3599–3606. ISSN: 1520-6882. DOI: 10.1021/ac2033084.

Zhang, Jiajie, Kassian Kobert, Tomáš Flouri, and Alexandros Stamatakis (Mar. 2014). "PEAR: a fast and accurate Illumina Paired-End reAd mergeR". en. In: *Bioinformatics* 30.5. Publisher: Oxford Academic, pp. 614–620. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btt593.

Zhang, Martin Jinye, Vasilis Ntranos, and David Tse (Feb. 2020). "Determining sequencing depth in a single-cell RNA-seq experiment". en. In: *Nature Communications* 11.1.

Number: 1 Publisher: Nature Publishing Group, p. 774. ISSN: 2041-1723. DOI: 10.1038/s41467-020-14482-y.

Zhang, Xiannian, Tianqi Li, Feng Liu, Yaqi Chen, Jiacheng Yao, Zeyao Li, Yanyi Huang, and Jianbin Wang (Jan. 2019). "Comparative Analysis of Droplet-Based Ultra-High-Throughput Single-Cell RNA-Seq Systems". en. In: *Molecular Cell* 73.1, 130–142.e5. ISSN: 10972765. DOI: 10.1016/j.molcel.2018.10.020.

Zheng, Grace X. Y. et al. (Jan. 2017). "Massively parallel digital transcriptional profiling of single cells". en. In: *Nature Communications* 8.1, pp. 1–12. ISSN: 2041-1723. DOI: 10.1038/ncomms14049.

Zhu, Jiang et al. (Apr. 2013). "Mining the antibodyome for HIV-1–neutralizing antibodies with next-generation sequencing and phylogenetic pairing of heavy/light chains". en. In: *Proceedings of the National Academy of Sciences* 110.16. ISBN: 9781219320110 Publisher: National Academy of Sciences Section: Biological Sciences, pp. 6470–6475. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1219320110.

Zhu, Y. Y., E. M. Machleder, A. Chenchik, R. Li, and P. D. Siebert (Apr. 2001). "Reverse transcriptase template switching: a SMART approach for full-length cDNA library construction". eng. In: *BioTechniques* 30.4, pp. 892–897. ISSN: 0736-6205. DOI: 10.2144/01304pf02.

Ziegenhain, Christoph et al. (Feb. 2017). "Comparative Analysis of Single-Cell RNA Sequencing Methods". en. In: *Molecular Cell* 65.4, 631–643.e4. ISSN: 1097-2765. DOI: 10.1016/j.molcel.2017.01.023.

Zilionis, Rapolas, Juozas Nainys, Adrian Veres, Virginia Savova, David Zemmour, Allon M. Klein, and Linas Mazutis (Jan. 2017). "Single-cell barcoding and sequencing using droplet microfluidics". en. In: *Nature Protocols* 12.1. Number: 1 Publisher: Nature Publishing Group, pp. 44–73. ISSN: 1750-2799. DOI: 10.1038/nprot.2016.154.

Zinchenko, Anastasia, Sean R. A. Devenish, Balint Kintses, Pierre-Yves Colin, Martin Fischlechner, and Florian Hollfelder (Mar. 2014). "One in a Million: Flow Cytometric Sorting of Single Cell-Lysate Assays in Monodisperse Picolitre Double Emulsion Droplets for Directed Evolution". In: *Analytical Chemistry* 86.5. Publisher: American Chemical Society, pp. 2526–2533. ISSN: 0003-2700. DOI: 10.1021/ac403585p.