THE UNIVERSITY OF CHICAGO

HIGH-THROUGHPUT MICROFLUIDIC PLATFORM FOR ULTRASENSITIVE SINGLE-CELL PROTEIN ANALYSES

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE PRITZKER SCHOOL OF MOLECULAR ENGINEERING

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

BY

JING LIN

CHICAGO, ILLINOIS

DECEMBER 2021

Table of Contents

List of Figures

List of Tables

Abstract

In the wake of heterogeneous nature of most biological systems, single-cell analysis has gained enormous importance and popularity in the field of biology, medicine and even microbiology. Traditional population-averaged analysis offers good information on the system level, however at the price of missing out cell-to-cell variations, which are critical information for building disease models, understanding drug resistance, or identifying cancer stem cells. Though single-cell transcriptomic technologies are well established, single-cell proteomic platforms seem to be lagging behind. This is mostly due to the low-level protein quantities present in single cells and the difficulty to amplify protein molecules for end-point detection. In this thesis, I will report microfluidic solutions for ultra-sensitive single-cell endogenous protein and enzymatic activities analysis. I will focus on the design principles of the microfluidic platforms, working pipeline validation and biological application demonstrations.

Firstly, I will demonstrate an ultrasensitive single-cell protein quantification platform based on a valve-based microfluidic chip designed and fabricated in house. I will show how the microfluidic platform improves the digital proximity ligation assay (digital PLA) performance, thus enabling quantification of rare protein species in single cells. Then, the assay will be extended to simultaneously quantify protein and mRNA copy numbers from the same single cell.

Then, the same microfluidic platform has been adapted to profile multiplexed enzymatic activities in single cells. By combining family-wide chemical probes and PLA, the assay - named single-cell activity-dependent proximity ligation (scADPL) – quantifies active enzymatic molecules of targeted protein species (instead of total molecules) in single cells. The proof-of-concept is then demonstrated with 6-plex measurements in cancer cell lines, and the assay has been applied to characterize cancer organoids derived from various patient samples.

Acknowledgements

First and foremost, I would like to thank my supervisor Dr. Savaş Tay for his invaluable guidance and tremendous support during my degree pursuit. Dr. Tay provided a very open research environment where I had the independence and freedom to think through and try out my hypotheses. And whenever I encountered bottlenecks, Dr. Tay always provided sharp insights that brought fruitful inspirations and helped me move forward effectively. Working with Dr. Tay sharpened my thinking and brought my work to a higher level.

I would also like to express my gratitude to the Tay Lab crews back when we were in Switzerland, for Drs. Ce Zhang and Michael Junkin to teach me everything about microfluidics, Dr. Christian Jordi to guide me through the digital PLA project. The weekly cross-border squash gatherings with Christian Jordi and Tino Frank were truly good memories. I would also like to thank my friends: Ce, Hsiung-Lin, Trang, Ruixia and Wei for making Basel feel like home.

My gratitude further extends to my Tay Lab colleagues at the University of Chicago. I would like to deeply thank Drs. Fatih Abasiyanik, Minjun Son, Luke Vistain, Nir Drayman and Hoang Van Phan for their unreserved sharing whenever I needed to discuss my projects with them for inspirations. I was also very fortunate to unite with my college roommate Judy Shi and get to know many new friends in Chicago: Xiao Liu, Yi Liao, Manman Shi, who helped me settle down in Chicago and get through difficult times. I will definitely miss the Friday food hunt scout: Van, Parthiv, Nicky and more, we created a lot of happy memories.

Lastly, I would like to thank my collaborators at the University of Chicago, Dr. Ray Moellering, Dr. Gang Li, and Kavya Pillai for their insightful discussions and devoted

contributions. It was a great honor to work with you all and I have learnt a great deal from you to become a better scientist.

Chapter 1

Introduction

1.1 The emergence of single-cell studies

Even cells stemmed from an isogenic population could demonstrate significant cell-to-cell heterogeneity, due to the intrinsic noise associated with gene expressions [1, 2]. Therefore, if we only study at the population level, the heterogeneous characteristics hidden in rare subpopulations will be obscured (**Fig. 1-1**). This phenomenon is especially well-known in the context of primary human carcinoma [3-8] and embryogenesis [9]. In the early 2000s, there was a boom of research works dedicated to understand single-cell heterogeneity in other biological systems as well, such as to characterize the bacteria population used in a fermentation process with Raman spectroscopy [10]; analyze single-cell transcripts during embryonic pancreatic development [11]; and investigate how external and internal noises associated with gene expressions and cellular signal processing determine the cell fates [12].

The early single-cell analyses were limited by the techniques available, such that they could only handle a small number of single cells, a small subpopulation or a limited panel of genetic markers. The isolation of single cells for downstream genetic or transcriptional analyses usually involved manual cell picking with a mouth pipette and a drawn glass needle into a PCR tube [13, 14], which is labor intensive and low-throughput. On the other hand, microscopy was also leveraged to image a small panel of transcripts to differentiate *C. elegans* lineages [15], and gene expression dynamics with spatio-temporal resolution [2, 16]. Although microscopy approaches enable simultaneous profiling of a relatively larger number of single cells, it requires pre-selection

of target gene/transcript panel, demands tedious transfection and experimental procedures and only provides approximate quantitative information based on fluorescence readout.

As the demand for better single-cell analysis techniques is increasingly apparent, various novel technologies have been developed that revolutionized the field. In the next sections, I will discuss more on the past innovations in the field of single-cell transcriptomics and proteomics.

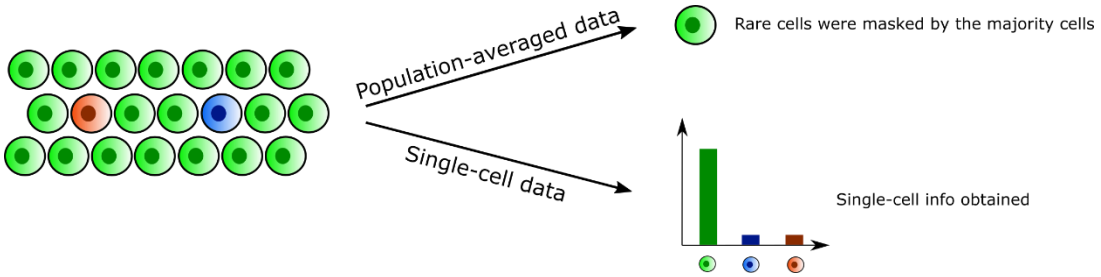

Figure 1-1: Schematic depicting the significance of single cell studies. In heterogeneous population, rare cells are often masked by the majority cell type, thus not showing up in the population-averaged data. In this case, single-cell study is useful to preserve the rare cell population information.

1.2 Single-cell transcriptomics

Transcriptomes (messenger RNAs - mRNAs) act as intermediate messengers between genomes (DNA) and proteins, thus could provide indicative information of the complete gene expression dynamics – which genes are actively being transcribed and which are silenced. They provide a comprehensive understanding of how the cell defines its type, responds to stimulus, or is healthy or diseased (**Fig. 1-2**). Transcriptomics are the gold standards to probe the cellular status, because of its ease to be amplified and numerous available readout techniques, which include reverse-transcriptase polymerase chain reaction (RT-PCR) [17, 18], microarrays [19, 20] and next-generation sequencing (NGS) [21-25].

Cell expression by linear amplification and sequencing (CEL-seq) [26] and switching mechanism at the 5’ end of the RNA transcript (SMART-seq) [21] were firstly commercialized with Fluidigm C1 system, which allows profiling of up to 96 single cells [27]. The single-cell

sorting step is performed on the C1 system, but also compatible with fluorescence-activated cell sorting (FACS) into 96 well plates. The procedure usually involves the capture of mRNA molecules on the poly-A tail, barcoding single cells, library preparation and library-pooling for sequencing. The major limitations of these techniques are tedious and low-throughput cell sorting and barcoding processes, which restrict the technique from profiling large number of single cells per run.

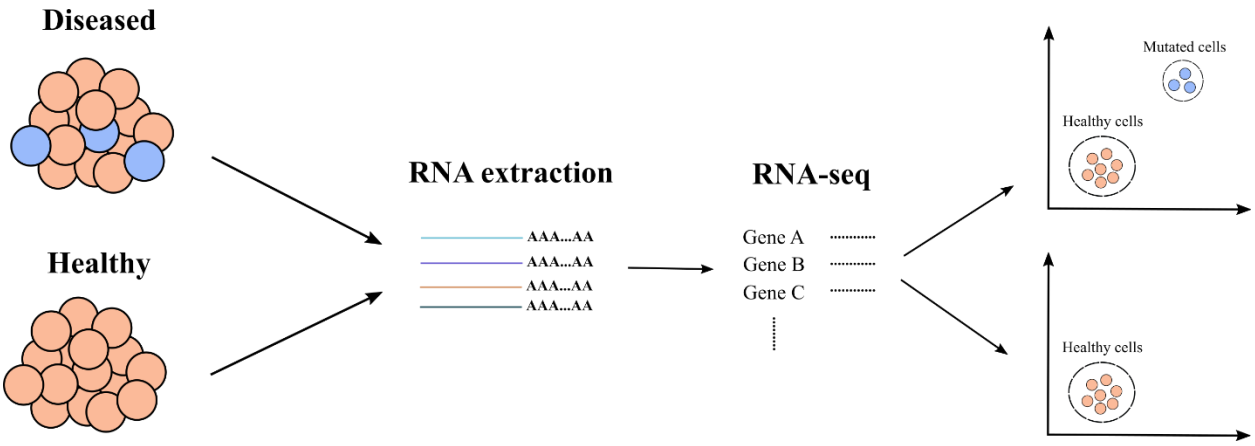

Figure 1-2: Schematics depicting the application of single-cell RNA-sequencing (scRNA-seq). It differentiates healthy and diseased tissues by spotting aberrant gene expression profiles associated with rare subpopulations.

Single-cell RNA-sequencing (scRNA-seq) became even more widely used after the innovation of high-throughput platforms, which enable the isolation of thousands of single cells in water-in-oil droplets and tagging the single cells with beads carrying cell-specific barcodes [24, 25]. This innovation requires relatively simple equipment setup, thus has greatly improved the applicability for any labs without microfluidic knowledge to conduct such experiments in house. Since then, many works were dedicated to improving the efficiency and performance of such platforms, such as single-cell capture rate, coverage, reads per gene and etc. [28-30]

1.3 Single-cell proteomics

1.3.1 The importance of single-cell proteomics

While single-cell transcriptomic sequencing is currently dominating the field due to its well-established pipelines [24, 25], it still has limitations. Transcripts (mRNAs), despite being easily accessible for sequencing, can only poorly represent the quantities of their respective encoded proteins, especially on single-cell level [31-33]. This poor mRNA-protein correlation might be due to different half-lives of mRNA and protein molecules, or each mRNA molecule is not translated equally. As proteins are the molecules that carry out the actual cellular functions, it is imperative to directly profile the cellular proteins instead of inferring from the mRNA data.

Despite its importance, the development of single-cell proteomic technology has been lagging. Early works utilized fluorescent proteins (such as GFP) to tag a single protein species for either dynamic gene expression monitoring (with time-lapse fluorescent microscopy) [34] or sorting (with flow cytometry) [35]; or tag the entire proteome to study the role of noise in overall protein expression regulation [36, 37]. However, fluorescence-based readout is limited by the number of channels available for high-level multiplexing and labor-intensive transfection procedure. On the other hand, antibody-based assays such as digital ELISA [38] and Luminex [39], enable targeted protein assays with high specificity and multiplexing capability. Antibody-based epitope profiling has also been integrated with high-throughput single-cell transcriptomic sequencing to allow single-cell multiomic analysis [40, 41]. For untargeted single-cell proteomic profiling, early prototype of single-cell mass spectrometry has been demonstrated to work with ~8 single cells with multiplexed 10 targets [42]. Although single-cell mass spectrometry is the future of single-cell proteomics, it still needs to overcome significant technical challenges before it can be widely implemented.

1.3.2 Proximity ligation-based single-cell proteomic technologies

The immunoassays mentioned in the previous section, which rely on a single readout antibody, suffer from high background noise due to unspecific bindings of the antibodies. The high background noise is especially problematic for single cells, in which case protein abundances tend to be low. To tackle this issue, proximity ligation assay (PLA) and its close relative proximity extension assay (PEA) were developed [43, 44]. PLA utilizes two readout oligonucleotides-conjugated antibodies (instead of one), which target at adjacent epitopes on the same protein molecule, such that ligation only happens between oligonucleotides that are on adjacent antibodies with the hybridization of universal complimentary splint oligos (**Fig. 1-3**). PLA provides the following advantages over conventional immunoassays: (1) it identifies a single protein molecule with a pair of antibodies that target at adjacent epitopes, thus reducing background noise from unspecific binding event of a single antibody; (2) convert protein signal to DNA signal, which can be easily amplified and quantified by polymerase chain reaction (PCR) or NGS.

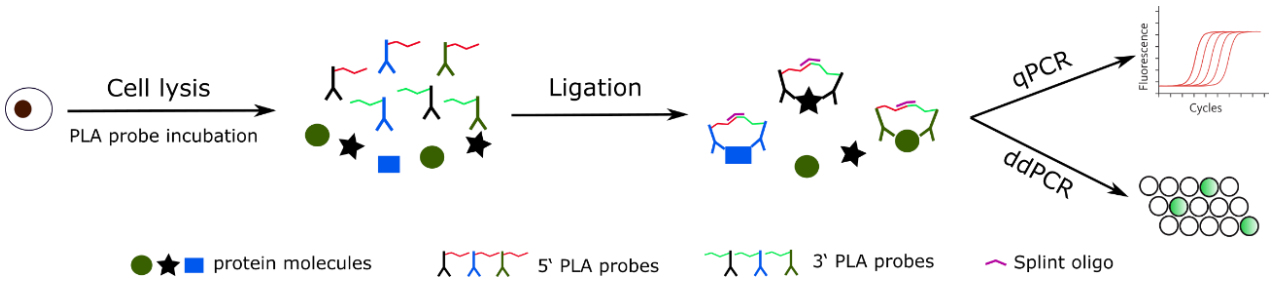

Figure 1-3: Schematic of PLA workflow. Single cells are isolated and lysed, cellular proteins are incubated with PLA probes, then ligation occurs for proximal oligonucleotides. The final PLA product can be either read out with qPCR (regular PLA) or ddPCR (digital PLA).

1.3.3 Digital PLA

PLA is read out with quantitative PCR (qPCR), which is prone to amplification efficiency variations among samples. Thus, our group developed a new technique, named digital PLA, that combines PLA with droplet digital PLA (ddPCR) (**Fig. 1-3**) [32]. Droplet digital PCR works by

partitioning amplicons (PLA products) into water-in-oil emulsions, which follows Poisson distribution, meaning most emulsion droplets are either empty or containing a single amplicon molecule. Thus, if the droplet is identified as positive (the presence of fluorescent signal) after PCR, it can be counted as one strand of amplicon in the initial sample, and the sample concentration can be calculated in an absolute manner [45]. Therefore, an emulsion droplet is either counted as positive (1) or negative (0) based on the presence or absence of the fluorescent signal after PCR, hence the name "digital PCR". Given the "digital" nature, ddPCR is immune to PCR amplification efficiency biases, and also able to resolve differences smaller than fold changes [46]. With these conferred advantages, digital PLA offers better sensitivity and resolution in single-cell protein profiling than regular PLA.

1.3.4 Activity-dependent proximity ligation assay

As so much effort is devoted for single-cell proteomics, it is however noteworthy that not all protein molecules are in active folding state at a specific time frame. Thus, quantifying the total number of protein molecules regardless their activities does not necessarily always lead to meaningful biological observations. Therefore, directly quantifying the active cellular proteins could be essential in understanding cellular dynamics, especially in cancers. A novel technique, named activity-dependent proximity ligation (ADPL) assay, combines family-wide chemical probes and PLA to probe the activity of a single protein species in situ [47] and in solution [48] (**Fig. 1-4**). Conventional protein activity probing involves the use of family-wide chemical probes, which target at the active sites of protein molecules, such as fluorophosphonate-biotin (FP-Bio) for active serine hydrolase enzymes. However, the conventional technique only quantifies the activities of the entire protein family but cannot resolve the activities of proteins belonging to the same family. To tackle this limitation, ADPL combines family-wide chemical probes with PLA,

in which one PLA probe (oligo-antibody conjugates) targeting at the protein species of interest and the other PLA probe targeting at streptavidin (cross-linked with the family-wide chemical probes, **Fig. 1-4**). The ADPL technology has been demonstrated with serine hydrolase and cysteine protease enzymes.

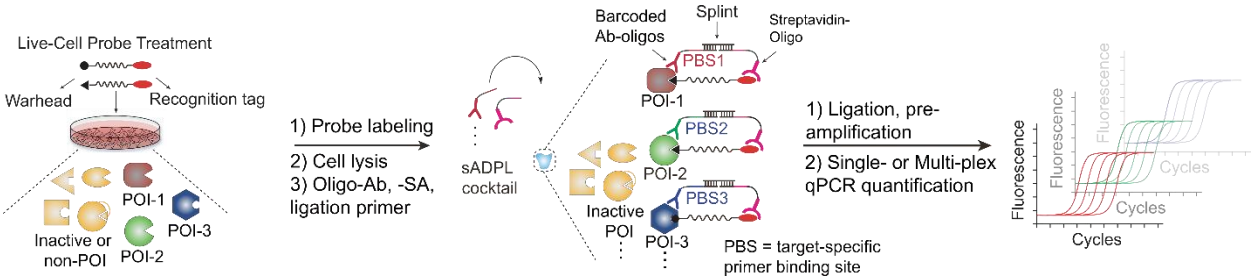

Figure 1-4: Schematics of bulk soluble ADPL workflow. Cells are treated with active protein-tagging probes; after cell lysis, tagged active proteins are hybridized with a pair of antibody probes with conjugated oligonucleotides; then the proximal oligos are ligated and read out with qPCR.

1.4 Microfluidic platforms for single-cell studies

As the term tells, microfluidics are platforms that manipulate small liquid volumes (in the range of picoliter to nanoliter) with micrometer-scale device features. Microfluidics confer numerous advantages comparing to conventional bench-top experimentation methods: firstly because of the small sample volume required by microfluidic platforms, the amount of reagents/samples consumption is significantly less, thus could enormously reduce experimental costs; secondly, integrated microfluidic platforms can accommodate much more experimental conditions and perform multi-dimensional analyses on the same cell in a simple setting; thirdly, the on-chip liquid manipulation is more precise than pipetting, thus yielding more consistent results; lastly, microfluidic platforms can potentially be automated, thus is much less laborious. There is also another unique characteristic of microfluidics that makes it highly desirable. Due to the micrometer-scale fluidic channels, the fluidic dynamics on microfluidic chips follow Laminar flow characteristics [49]. In contrast to turbulent flow, Laminar flow is characterized as two joining

flow streams retaining their paths with minimum lateral mixing, while mixing is only achievable through diffusion across the interface over time. Because of this unique property, microfluidics are widely used for biological studies in a controllable chemical gradient setting [50-55], which cannot be easily achieved by other platforms.

As microfluidic systems have micrometer-scale features that are in the same scale of a single mammalian cell (~10 μm), they are therefore ideal platforms for single-cell analyses for the following reasons. Firstly, single cells can be easily sorted on microfluidic chips, thus alleviating the needs for complicated sorting apparatus such as FACS [56-58], as well as allowing downstream analyses on the same integrated platform. Secondly, single cells can be isolated in individual chambers that are separated by microfluidic solenoid valves [59] or water-in-oil droplets [24], in which each single cell can be manipulated individually: receiving different stimuli, interacting with another single cell, be retrieved selectively for downstream analyses and etc. [60, 61] Lastly, microfluidic platforms are mostly consisted of materials such as polydimethylsiloxane (PDMS) and glass, which are biocompatible and optically permissible that allow easy microscopic imaging.

As has been briefly mentioned earlier, there are various types of microfluidics: valve-based, droplet-based, paper-based, digital microfluidics and etc. Among those, valve-based and droplet-based microfluidics are most widely used for single-cell analyses. Therefore, I will discuss more about these two types in the following sections.

1.4.1 Valve-based microfluidics

To manipulate single cells independently on microfluidic platforms, it is critical to ensure complete closing of chambers with no crosstalk among them. To achieve this, microfluidic valves were invented [62]. The microfluidic valve works as shown in **Fig. 1-5**: the valve consists of two layers of PDMS devices fabricated with soft lithography, normally the bottom layer is the “thin

layer" that fills with water (we call this assemble "push-up microfluidic valve", if the thin layer is placed on top, it is called "push-down microfluidic valve"); once the channel filled with water is pressurized, it can deform the thin layer such that it will completely block the channel on the other side, thus forming a closed valve (**Fig. 1-5b**); once the pressure is released, the valve will be open and flow can resume in the channel (**Fig. 1-5a**).

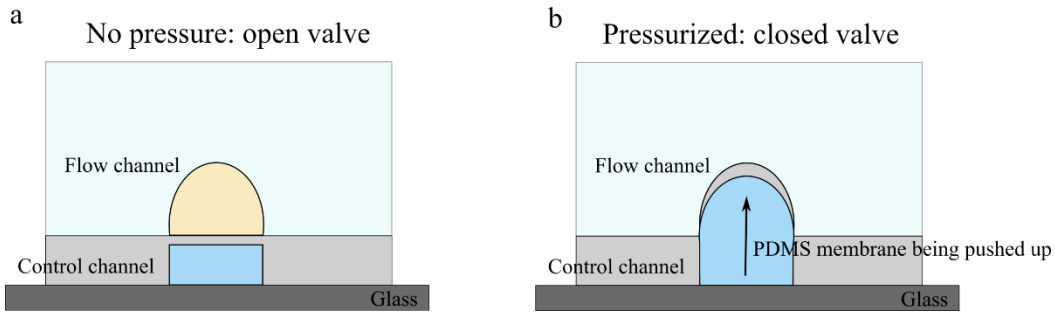

Figure 1-5: A schematic showing the cross-section of a microfluidic valve working mechanism. The control channel and flow channel are separated by a thin PDMS membrane. While control channel is filled with water, the flow channel handles experimental reagents/samples. (**a**) When there is no pressure applied to control channel, the valve is open and flow continues; (**b**) when the control channel is pressurized, the membrane gets pushed up and the flow channel is thus blocked.

Valve-based microfluidics are extensively used for applications where precise fluidic control is needed. This includes precise spatiotemporal reagents delivery to specific chambers [63, 64], reagent delivery with peristaltic pumps for highly controllable volumes [62, 65], and time-course study of single-cell secretions [66, 67]. Although valve-based microfluidics confer significant advantages with precise liquid delivery and manipulation, they are however difficult to fabricate and operate, which limit their wide uses in non-microfluidic labs.

1.4.2 Droplet-based microfluidics

Instead of trapping cells in a microfluidic chamber, there is another type of microfluidics that performs reactions inside droplets (for biological applications, normally water-in-oil emulsions are used) [68]. Droplets are generated by flowing two liquid phases towards a flow-

focusing junction with fine-tuned flow rate parameters, such that the continuous phase liquid is "pinching" the discrete phase liquid into numerous independent droplets of approximately the same size. The droplets therefore contain the discrete phase and are separated from the other droplets by the continuous phase (**Fig. 1-6**).

Droplet microfluidics have less precise control over each individual droplet, as comparing to valve-based microfluidics to each chamber, however, droplet microfluidics confer apparent advantage of enabling high-throughput assays by generating thousands of droplets every second. It is more challenging to achieve such high throughput with valve-based microfluidics, as more chambers are demanded to accommodate more samples (which exponentially increases the fabrication difficulties); on the contrary, droplet microfluidics are more scalable in that aspect. Besides being high-throughput, droplet microfluidics have comprehensive pipelines to perform various experimental procedures: add reagents into existing droplets (pico-injection) [69-71], merge droplets [72, 73], culture and sort droplets selectively based on their contents [74-76]. The comprehensive pipelines (and more to come in the near future) would allow most lab procedures to transfer into droplets, which could substantially improve the capability and throughputs of typical laboratories with much less reagent consumption. Currently, droplet microfluidics are mostly used in high-throughput single-cell genomic/transcriptomic sequencing, in which thousands of single cells are encapsulated and barcoded in droplets [24, 25].

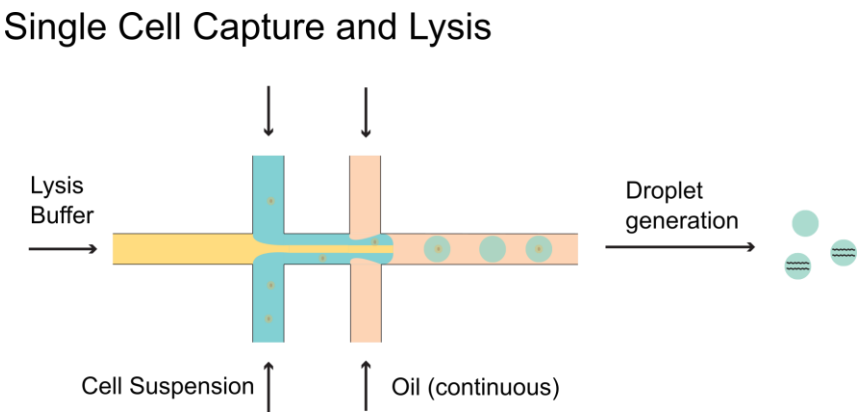

Figure 1-6: Schematic showing the mechanism of droplet formation. An example is how single cells are encapsulated and lysed in water-in-oil droplets, by flowing reagents (cell suspension and lysis buffer) and oil towards a flow focusing junction.

1.5 Thesis overview

In this thesis, I will report a novel valve-based microfluidic platform that is specifically designed to enable single-cell protein analyses with ultrahigh sensitivities. I will mainly focus on characterizing how the microfluidic platform improves the assay performances, and how the platform can be generally applied to other protein assays.

In Chapter 2, I will present the novel microfluidic platform, its design principles, considerations and the fabrication methods. I will then characterize how much the microfluidic platform improves digital PLA (μ-dPLA) as comparing to the assay performance without using the chip in terms of both sensitivity and assay efficiency (protein molecules/amplicon), by performing calibration curves with purified protein standards. Afterwards, μ-dPLA has been leveraged to quantify single-cell endogenous protein targets with low abundances (TNFR1), which is below the limit of detection without using the microfluidic platform. Lastly, the technique has been demonstrated to simultaneously measure protein and RNA molecules from the same single cells, thus providing an extra dimension of data into deciphering the cells [77].

In Chapter 3, I will integrate the same microfluidic platform presented in Chapter 2 with ADPL to profile multiplexed enzyme activities in single cells, a technique named scADPL. This is the first data that we know of to show protein activity profiles in single cells. I will firstly show proof-of-concept data, which demonstrate that scADPL can robustly differentiate treated and untreated (with the family-wide chemical probes) cell populations. Then, I will show the assay specificity by demonstrating that with a specific protein activity inhibitor, scADPL can consistently spot the inhibited activity while the other protein targets readings are unaffected. After that, scADPL has been leveraged to measure a panel of 6 cancer metastasis-related biomarkers to differentiate two cell lines of the same cancer type but different phenotypes (SKOV3 vs. OVCAR3). Lastly, scADPL has been applied to profile single cells dissociated from breast cancer organoids that were derived from a single patient sampled from various metastatic sites.

In Appendix, I will describe a project in which I developed high-throughput droplet-based microfluidic system that could enable transcription dynamics profiling in thousands of single cells (scKAS-seq). In this part, I will present the project rationale, microfluidic pipeline design principles and preliminary data.

1.6 Bibliography

1. Wang, D. and S. Bodovitz, *Single cell analysis: the new frontier in 'omics'.* Trends in biotechnology, 2010. **28**(6): p. 281-290.
2. Elowitz, M.B., et al., *Stochastic gene expression in a single cell.* Science, 2002. **297**(5584): p. 1183-1186.
3. Fujii, H., et al., *Genetic divergence in the clonal evolution of breast cancer.* Cancer Research, 1996. **56**(7): p. 1493-1497.
4. Macintosh, C.A., et al., *Precise microdissection of human prostate cancers reveals genotypic heterogeneity.* Cancer research, 1998. **58**(1): p. 23-28.
5. Heim, S., et al., *Cytogenetic polyclonality in tumors of the breast.* Cancer genetics and cytogenetics, 1997. **95**(1): p. 16-19.
6. Fisher, R., L. Pusztai, and C. Swanton, *Cancer heterogeneity: implications for targeted therapeutics.* British journal of cancer, 2013. **108**(3): p. 479-485.

7. Bhatia, S., et al., *The challenges posed by cancer heterogeneity.* Nature biotechnology, 2012. **30**(7): p. 604-610.

8. Klco, J.M., et al., *Functional heterogeneity of genetically defined subclones in acute myeloid leukemia.* Cancer cell, 2014. **25**(3): p. 379-392.

9. Chen, Q., et al., *Tracing the origin of heterogeneity and symmetry breaking in the early mammalian embryo.* Nature communications, 2018. **9**(1): p. 1-11.

10. Schuster, K.C., E. Urlaub, and J. Gapes, *Single-cell analysis of bacteria by Raman microscopy: spectral information on the chemical composition of cells and on the heterogeneity in a culture.* Journal of Microbiological Methods, 2000. **42**(1): p. 29-38.

11. Chiang, M.-K. and D.A. Melton, *Single-cell transcript analysis of pancreas development.* Developmental cell, 2003. **4**(3): p. 383-393.

12. Spiller, D.G., et al., *Measurement of single-cell dynamics.* Nature, 2010. **465**(7299): p. 736-745.

13. Trimarchi, J.M., et al., *Molecular heterogeneity of developing retinal ganglion and amacrine cells revealed through single cell gene expression profiling.* Journal of Comparative Neurology, 2007. **502**(6): p. 1047-1065.

14. Dixon, A.K., et al., *Gene-expression analysis at the single-cell level.* Trends in pharmacological sciences, 2000. **21**(2): p. 65-70.

15. Liu, X., et al., *Analysis of cell fate from single-cell gene expression profiles in C. elegans.* Cell, 2009. **139**(3): p. 623-633.

16. Rosenfeld, N., et al., *Gene regulation at the single-cell level.* science, 2005. **307**(5717): p. 1962-1965.

17. Tiller, T., et al., *Efficient generation of monoclonal antibodies from single human B cells by single cell RT-PCR and expression vector cloning.* Journal of immunological methods, 2008. **329**(1-2): p. 112-124.

18. Wang, X. and B.D. Stollar, *Human immunoglobulin variable region gene analysis by single cell RT-PCR.* Journal of immunological methods, 2000. **244**(1-2): p. 217-225.

19. Moncada, R., et al., *Integrating microarray-based spatial transcriptomics and single-cell RNA-seq reveals tissue architecture in pancreatic ductal adenocarcinomas.* Nature biotechnology, 2020. **38**(3): p. 333-342.

20. Kurimoto, K., et al., *Global single-cell cDNA amplification to provide a template for representative high-density oligonucleotide microarray analysis.* Nature protocols, 2007. **2**(3): p. 739-752.

21. Ramsköld, D., et al., *Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells.* Nature biotechnology, 2012. **30**(8): p. 777-782.

22. Picelli, S., et al., *Smart-seq2 for sensitive full-length transcriptome profiling in single cells.* Nature methods, 2013. **10**(11): p. 1096-1098.

23. Hagemann-Jensen, M., et al., *Single-cell RNA counting at allele and isoform resolution using Smart-seq3.* Nature Biotechnology, 2020. **38**(6): p. 708-714.

24. Macosko, E.Z., et al., *Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets.* Cell, 2015. **161**(5): p. 1202-1214.

25. Klein, A.M., et al., *Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells.* Cell, 2015. **161**(5): p. 1187-1201.

26. Hashimshony, T., et al., *CEL-Seq: single-cell RNA-Seq by multiplexed linear amplification.* Cell reports, 2012. **2**(3): p. 666-673.

27. Ziegenhain, C., et al., *Comparative analysis of single-cell RNA sequencing methods.* Molecular cell, 2017. **65**(4): p. 631-643. e4.

28. Cao, J., et al., *Comprehensive single-cell transcriptional profiling of a multicellular organism.* Science, 2017. **357**(6352): p. 661-667.

29. Saliba, A.-E., et al., *Single-cell RNA-seq: advances and future challenges.* Nucleic acids research, 2014. **42**(14): p. 8845-8860.

30. Rodriques, S.G., et al., *Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution.* Science, 2019. **363**(6434): p. 1463-1467.

31. Taniguchi, Y., et al., *Quantifying E. coli proteome and transcriptome with single-molecule sensitivity in single cells.* science, 2010. **329**(5991): p. 533-538.

32. Albayrak, C., et al., *Digital quantification of proteins and mRNA in single mammalian cells.* Molecular cell, 2016. **61**(6): p. 914-924.

33. Darmanis, S., et al., *Simultaneous multiplexed measurement of RNA and proteins in single cells.* Cell reports, 2016. **14**(2): p. 380-389.

34. Muzzey, D. and A. van Oudenaarden, *Quantitative time-lapse fluorescence microscopy in single cells.* Annual Review of Cell and Developmental, 2009. **25**: p. 301-327.

35. Anderson, M., et al., *Simultaneous fluorescence-activated cell sorter analysis of two distinct transcriptional elements within a single cell using engineered green fluorescent proteins.* Proceedings of the National Academy of Sciences, 1996. **93**(16): p. 8508-8511.

36. Newman, J.R., et al., *Single-cell proteomic analysis of S. cerevisiae reveals the architecture of biological noise.* Nature, 2006. **441**(7095): p. 840-846.

37. Cai, L., N. Friedman, and X.S. Xie, *Stochastic protein expression in individual cells at the single molecule level.* Nature, 2006. **440**(7082): p. 358-362.

38. Rissin, D.M., et al., *Single-molecule enzyme-linked immunosorbent assay detects serum proteins at subfemtomolar concentrations.* Nature biotechnology, 2010. **28**(6): p. 595-599.

39. Chowdhury, F., A. Williams, and P. Johnson, *Validation and comparison of two multiplex technologies, Luminex® and Mesoscale Discovery, for human cytokine profiling.* Journal of immunological methods, 2009. **340**(1): p. 55-64.

40. Stoeckius, M., et al., *Simultaneous epitope and transcriptome measurement in single cells.* Nature methods, 2017. **14**(9): p. 865.

41. Peterson, V.M., et al., *Multiplexed quantification of proteins and transcripts in single cells.* Nature biotechnology, 2017. **35**(10): p. 936.

42. Budnik, B., et al., *SCoPE-MS: mass spectrometry of single mammalian cells quantifies proteome heterogeneity during cell differentiation.* Genome biology, 2018. **19**(1): p. 161.

43. Fredriksson, S., et al., *Protein detection using proximity-dependent DNA ligation assays.* Nature biotechnology, 2002. **20**(5): p. 473.

44. Söderberg, O., et al., *Direct observation of individual endogenous protein complexes in situ by proximity ligation.* Nature methods, 2006. **3**(12): p. 995-1000.

45. Hindson, B.J., et al., *High-throughput droplet digital PCR system for absolute quantitation of DNA copy number.* Analytical chemistry, 2011. **83**(22): p. 8604-8610.

46. Fan, H.C., et al., *Non-invasive prenatal measurement of the fetal genome.* Nature, 2012. **487**(7407): p. 320-324.

47. Li, G., et al., *An activity-dependent proximity ligation platform for spatially resolved quantification of active enzymes in single cells.* Nature communications, 2017. **8**(1): p. 1-12.

48. Li, G., et al., *Ultrasensitive, multiplexed chemoproteomic profiling with soluble activity-dependent proximity ligation.* Proceedings of the National Academy of Sciences, 2019. **116**(43): p. 21493-21500.

49. Yin, H. and D. Marshall, *Microfluidics for single cell analysis.* Current opinion in biotechnology, 2012. **23**(1): p. 110-119.

50. Jeon, N.L., et al., *Generation of solution and surface gradients using microfluidic systems.* Langmuir, 2000. **16**(22): p. 8311-8316.

51. Kim, S., H.J. Kim, and N.L. Jeon, *Biological applications of microfluidic gradient devices.* Integrative Biology, 2010. **2**(11-12): p. 584-603.

52. Irimia, D., D.A. Geba, and M. Toner, *Universal microfluidic gradient generator.* Analytical chemistry, 2006. **78**(10): p. 3472-3477.

53. Chung, B.G., et al., *Human neural stem cell growth and differentiation in a gradient-generating microfluidic device.* Lab on a Chip, 2005. **5**(4): p. 401-406.

54. Chung, B.G. and J. Choo, *Microfluidic gradient platforms for controlling cellular behavior.* Electrophoresis, 2010. **31**(18): p. 3014-3027.

55. Frank, T. and S. Tay, *Flow-switching allows independently programmable, extremely stable, high-throughput diffusion-based gradients.* Lab on a Chip, 2013. **13**(7): p. 1273-1281.

56. Chung, K., et al., *Imaging single-cell signaling dynamics with a deterministic high-density single-cell trap array.* Analytical chemistry, 2011. **83**(18): p. 7044-7052.

57. Kobel, S., et al., *Optimization of microfluidic single cell trapping for long-term on-chip culture.* Lab on a Chip, 2010. **10**(7): p. 857-863.

58. Jin, D., et al., *A microfluidic device enabling high-efficiency single cell trapping.* Biomicrofluidics, 2015. **9**(1): p. 014101.

59. Gómez-Sjöberg, R., et al., *Versatile, fully automated, microfluidic cell culture system.* Analytical chemistry, 2007. **79**(22): p. 8557-8563.

60. Mehling, M., et al., *Real-time tracking, retrieval and gene expression analysis of migrating human T cells.* Lab on a Chip, 2015. **15**(5): p. 1276-1283.

61. Dettinger, P., et al., *Automated microfluidic system for dynamic stimulation and tracking of single cells.* Analytical chemistry, 2018. **90**(18): p. 10695-10700.

62. Unger, M.A., et al., *Monolithic microfabricated valves and pumps by multilayer soft lithography.* Science, 2000. **288**(5463): p. 113-116.

63. Tay, S., et al., *Single-cell NF-κB dynamics reveal digital activation and analogue information processing.* Nature, 2010. **466**(7303): p. 267-271.

64. Zhang, C., et al., *Ultra-multiplexed analysis of single-cell dynamics reveals logic rules in differentiation.* Science advances, 2019. **5**(4): p. eaav7959.

65. Chou, H.-P., M.A. Unger, and S.R. Quake, *A microfabricated rotary pump.* Biomedical Microdevices, 2001. **3**(4): p. 323-330.

66. Kaestli, A.J., M. Junkin, and S. Tay, *Integrated platform for cell culture and dynamic quantification of cell secretion.* Lab on a Chip, 2017. **17**(23): p. 4124-4133.

67. Junkin, M., et al., *High-content quantification of single-cell immune dynamics.* Cell reports, 2016. **15**(2): p. 411-422.

68. Teh, S.-Y., et al., *Droplet microfluidics.* Lab on a Chip, 2008. **8**(2): p. 198-220.

69. Abate, A.R., et al., *High-throughput injection with microfluidics using picoinjectors.* Proceedings of the National Academy of Sciences, 2010. **107**(45): p. 19163-19166.

70. Eastburn, D.J., A. Sciambi, and A.R. Abate, *Picoinjection enables digital detection of RNA with droplet rt-PCR.* PloS one, 2013. **8**(4): p. e62961.

71. O'Donovan, B., D.J. Eastburn, and A.R. Abate, *Electrode-free picoinjection of microfluidic drops.* Lab on a Chip, 2012. **12**(20): p. 4029-4032.

72. Lan, F., et al., *Single-cell genome sequencing at ultra-high-throughput with microfluidic droplet barcoding.* Nature biotechnology, 2017. **35**(7): p. 640-646.

73. Shahi, P., et al., *Abseq: Ultrahigh-throughput single cell protein profiling with droplet microfluidic barcoding.* Scientific reports, 2017. **7**(1): p. 1-12.

74. Watterson, W.J., et al., *Droplet-based high-throughput cultivation for accurate screening of antibiotic resistant gut microbes.* Elife, 2020. **9**: p. e56998.

75. Baret, J.-C., et al., *Fluorescence-activated droplet sorting (FADS): efficient microfluidic cell sorting based on enzymatic activity.* Lab on a Chip, 2009. **9**(13): p. 1850-1858.

76. Xi, H.-D., et al., *Active droplet sorting in microfluidics: a review.* Lab on a Chip, 2017. **17**(5): p. 751-771.

77. Lin, J., et al., *Ultra-sensitive digital quantification of proteins and mRNA in single cells.* Nature communications, 2019. **10**(1): p. 1-10.

Chapter 2

Ultrasensitive digital quantification of protein and mRNA in single cells

2.1 Introduction

Despite significant advances in single-cell analysis methods [1-4], techniques for high-resolution protein copy number quantification are lagging behind. Current techniques such as mass spectrometry (MS) generally lack the sensitivity to detect the small amounts of proteins present in individual cells [5, 6]; however, recent developments in MS are promising in detecting proteins in single cells [7]. Flow cytometry and mass cytometry (e.g., CyTOF) can detect proteins in single cells, but developing sample standards for quantification is challenging [8]. Protein immunoassays such as ELISA or enzyme-linked immunospot assay can detect proteins from single cells but lack the sensitivity for accurate quantification. CITE-seq (cellular indexing of transcriptomes and epitopes by sequencing) [9] and REAP-seq (RNA expression and protein sequencing) [10] are microfluidic- and sequencing-based technologies that quantify single-cell transcriptome and epitome simultaneously in a high-throughput manner; however, they are not compatible with intracellular proteins. These methods also have limited resolution due to the high background from non-specific binding of antibodies.

To overcome those limitations, digital proximity ligation assay (dPLA) was recently introduced. dPLA provides the ability of direct and digital measurements of protein and mRNA copy numbers in single mammalian cells [11]. In dPLA, digital PCR (dPCR) is used to quantify proteins detected with a pair of oligonucleotide-tagged antibodies called PLA probes. Previously published PLA (or its close relative PEA) methods enable multiplexed simultaneous protein and mRNA measurement from single cells. Those methods claim to multiplex ~96 protein targets in a

single cell; however, their quantitative polymerase chain reaction (qPCR) readout limits the sensitivity of the measurements [12-14]. The use of the dPCR readout provides significantly improved resolution and limit of detection (LOD) [15, 16], which allows direct quantification of protein copy numbers in individual mammalian cells (qPCR and dPCR comparisons are shown in Methods and **Table 2-1** and **Fig. 2-1**) [11]. This advance, along with the use of dPCR for mRNA quantification, enables mathematical models of gene expression and protein translation in single mammalian cells [11]. However, currently dPLA is not able to quantify rare protein targets because of the large dilution factor introduced by sorting and lysing individual cells in micro-well plates. Further, the current dPLA method is not compatible with single-cell imaging, tracking, and chemical stimulation [11, 17]. These limitations hinder many applications of regular PLA and dPLA for single-cell analysis. Genshaft et al. [18] has presented a similar method to simultaneously profile protein (multiplexed ~27 targets) and mRNA in single cells on commercial microfluidic platforms. However, this method relies on qPCR readout and requires pre-amplification for both single-cell RNA and protein quantification, which can often create bias in the results and is not compatible with single-cell imaging, tracking, and chemical stimulation.

Here, we present an automated microfluidic device for dPLA measurements (μ-dPLA) that also combines live-cell imaging and chemical stimulation in the same platform. Most importantly, our system improves the detection limit and sensitivity of single-cell digital protein quantification by up to 55-fold, above an absolute limit 2277 copies of proteins per cell, and with detection efficiency (i.e., number of protein molecules converted to one DNA molecule in PLA, the detection efficiency is calculated as in Eq. (1) in Methods) of as few as 29 protein molecules, thus expanding the applicability of dPLA in single-cell studies dramatically. Our method confers the additional advantage of combining live-cell microscopy and dynamic single-cell stimulation with

digital protein and mRNA measurements for multi-parametric profiling, which enables studying protein dynamics in living cells and linking it to end-point protein and mRNA measurements. Furthermore, our approach significantly reduces reagent cost and labor, and improves overall reproducibility of single-cell measurements via automated cell and reagent handling.

Table 2-1: Two tables show *p*-values calculated from pairwise t-test between every 2 concentrations for qPCR (table above) and ddPCR (table below). The first row and column of both tables are TNFR1 purified protein concentrations with unit of ng ml^{-1}. The lower the *p*-values confers that the oligo concentration differences can be better resolved.

qPCR	1	1.05	1.1	1.15	1.2	1.25	1.3
1	N.A.	0.191	0.244	0.035	0.025	0.026	0.016
1.05	0.191	N.A.	0.852	0.094	0.01	0.012	0.003
1.1	0.244	0.852	N.A.	0.095	0.028	0.031	0.013
1.15	0.035	0.094	0.095	N.A.	0.427	0.499	0.084
1.2	0.025	0.01	0.028	0.427	N.A.	0.842	0.05
1.25	0.026	0.012	0.031	0.499	0.842	N.A.	0.043
1.3	0.016	0.003	0.013	0.084	0.05	0.043	N.A.

ddPCR	1	1.05	1.1	1.15	1.2	1.25	1.3
1	N.A.	0.508	0.06	0.01	0.009	0.00036	0.00051
1.05	0.508	N.A.	0.325	0.069	0.0098	0.0019	0.002
1.1	0.06	0.325	N.A.	0.054	0.03	0.0022	0.0032
1.15	0.01	0.069	0.054	N.A.	0.052	0.002	0.0031
1.2	0.009	0.0098	0.03	0.052	N.A.	0.177	0.236
1.25	0.00036	0.0019	0.0022	0.002	0.177	N.A.	0.74
1.3	0.00051	0.002	0.0032	0.0031	0.236	0.74	N.A.

We applied our technology to study the correlation between protein and mRNA copy numbers in individual mammalian cells and found higher degrees of mRNA-protein correlation compared to previous estimates [11]. We showed that the ability of accurately measuring protein abundance in lowly expressing cells – which were not accounted for in previous measurements due to limited sensitivity – allow accurate characterization of mRNA-protein correlation. Finally, we used combined live-cell imaging and digital protein quantification in single cells, a new capability provided by our technology, and studied virus infection dynamics of lung epithelial cells.

Live-cell imaging of herpes simplex virus 1 (HSV-1)-infected lung epithelial cells show that the expression of ICP4 – a major transcription factor that regulates hundreds of viral genes – is only partially correlated with the total viral protein counts in single cells, suggesting that not all the cells infected by HSV-1 are able to go past the point of immediate-early protein synthesis and go through abortive infection. Our results highlight the importance of high-sensitivity digital measurements in understanding the relationship between gene expression and protein production in single cells in different contexts.

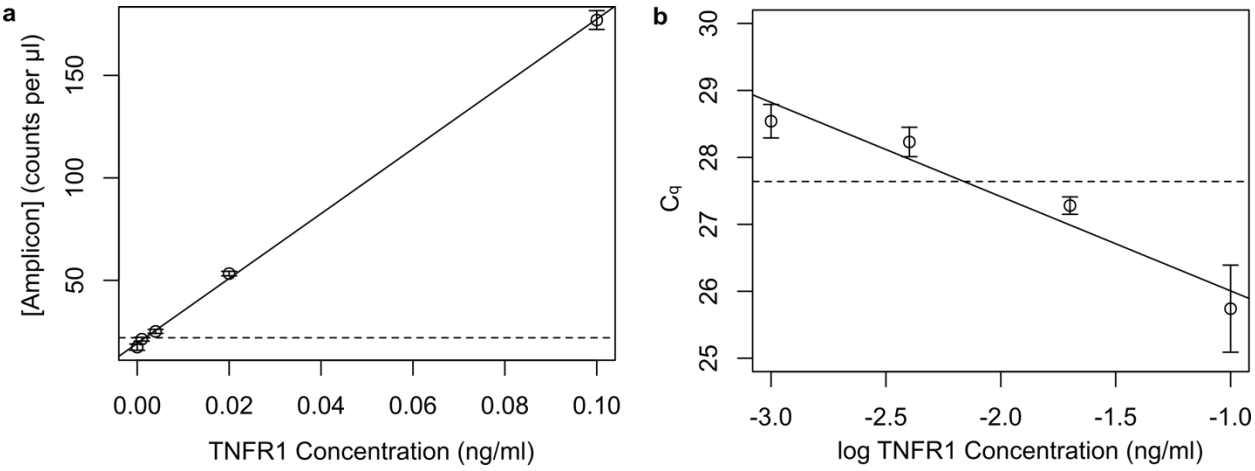

Figure 2-1: Side-by-side comparison of qPCR and ddPCR with TNFR1 calibration curve. (**a**) TNFR1 calibration curve readout with ddPCR, $y = 19.2726 + 1581.9x$, $R^2 = 0.9995$, the dashed line indicates the LOD = 1.7 pg ml^{-1}, (**b**) TNFR1 calibration curve readout with qPCR, $y = 24.599 – 1.407x$, $R^2 = 0.9369$, the dashed line indicates the LOD = 6.9 pg ml^{-1}. Error bars calculation is mentioned in the Methods section.

2.2 Results and discussions

2.2.1 Development of µ-dPLA system

In regular dPLA, cells are sorted and lysed in regular micro-wells, and then they are combined with PLA probes that bind to the target proteins. PLA probes consisting of oligonucleotide-functionalized antibodies recognize the target protein at two different epitopes, and oligonucleotides are then ligated with the help of a short piece of connector due to extreme proximity on the target protein. Droplet digital PCR (ddPCR) is used to quantify the hybridized

oligonucleotides, allowing direct counting of protein copy numbers in the sample. While regular dPLA can achieve LOD at the sub-femtomolar range, the large dilution of the single-cell contents into the micro-wells results in reduced sensitivity for practical applications. Therefore, dPLA is unable to quantify protein abundances below ~30,000 copies per cell, and rarely expressed proteins are missed altogether [11].

To improve the sensitivity of dPLA for profiling rare proteins in single cells, our approach was to sort and confine cells in nanoliter fluidic chambers at every step of the dPLA protocol. We developed an automated microfluidic device that captures and sorts individual cells into 7-nl chambers. This allowed us to dramatically reduce the dilution factor introduced from single cell sorting and lysis steps. Furthermore, we avoided sample loss by reducing the reaction volume, so that 100% of the single-cell sample was used in the ddPCR readout. In contrast, only 9% of the sample was used in regular dPLA, which results in 91% loss of signal [11]. Furthermore, the microfluidic system performs all assay steps automatically and performs many repeated fluidic manipulations with nanoliter precision. While the typical sampling error is at ~10% when manual pipetting is used, but our automated systems reduces these pipetting errors to <0.1%, resulting in dramatically improved technical reproducibility.

The integrated microfluidic device designed for µ-dPLA is capable of performing 144 parallel single-cell dPLA assays in an automated fashion (**Fig. 2-2**). The device traps and lyses single cells in a 7-nl chamber (**Fig. 2-2a**, chamber III). To further minimize dilution factor throughout the dPLA process, a series of microfluidic chambers were designed to precisely execute each assay step without sample loss. To avoid contamination between steps and enable automation, each chamber was separated by a set of microfluidic valves [19], which opens specific chambers at the programmed time. Chip automation was performed by controlling valve operation using a

Matlab program [17]. Measuring of precise reagents volumes was achieved by delivering reagents to empty chambers with predefined volumes, a process called dead-end filling. Wash channels were designed to prevent contamination of reagents between steps. The use of automated nanoliter fluid manipulation has the additional advantage of extremely reproducible pipetting, which reduces technical errors and variability, resulting in increased precision and sensitivity in protein and mRNA measurements.

2.2.2 Ultrasensitive digital protein measurement in single cells

To characterize the performance of our μ-dPLA device for protein quantification, we generated calibration curves with pure protein standards (**Fig. 2-3a, d and Fig. 2-4c**). These curves show the measured number of double-stranded DNA amplicons per microliter (ddPCR readout) against the total protein molecules in chamber III (**Fig. 2-2**). Table 2-2 shows the LOD and detection efficiency for endogenous human proteins CD147, TNFR1, and CSTB for both on- and off-chip (regular) dPLA. We found that μ-dPLA has up to 55 times better LOD and 25 times better detection efficiency than regular dPLA. This extremely high sensitivity enables quantification of very rare protein targets: we were able to quantify 2277, 2795, and 18,824 molecules of TNFR1, CD147, and CSTB protein individual cells, respectively. Our device was able to achieve detection efficiency as low as 33, 29, and 304 copies of TNFR1, CD147, and CSTB protein molecules, respectively. Moreover, our system achieved a linear dynamic range of 3 – 4 orders of magnitude for these proteins. To our knowledge, this is the highest demonstrated sensitivity to date for protein quantification in single mammalian cells.

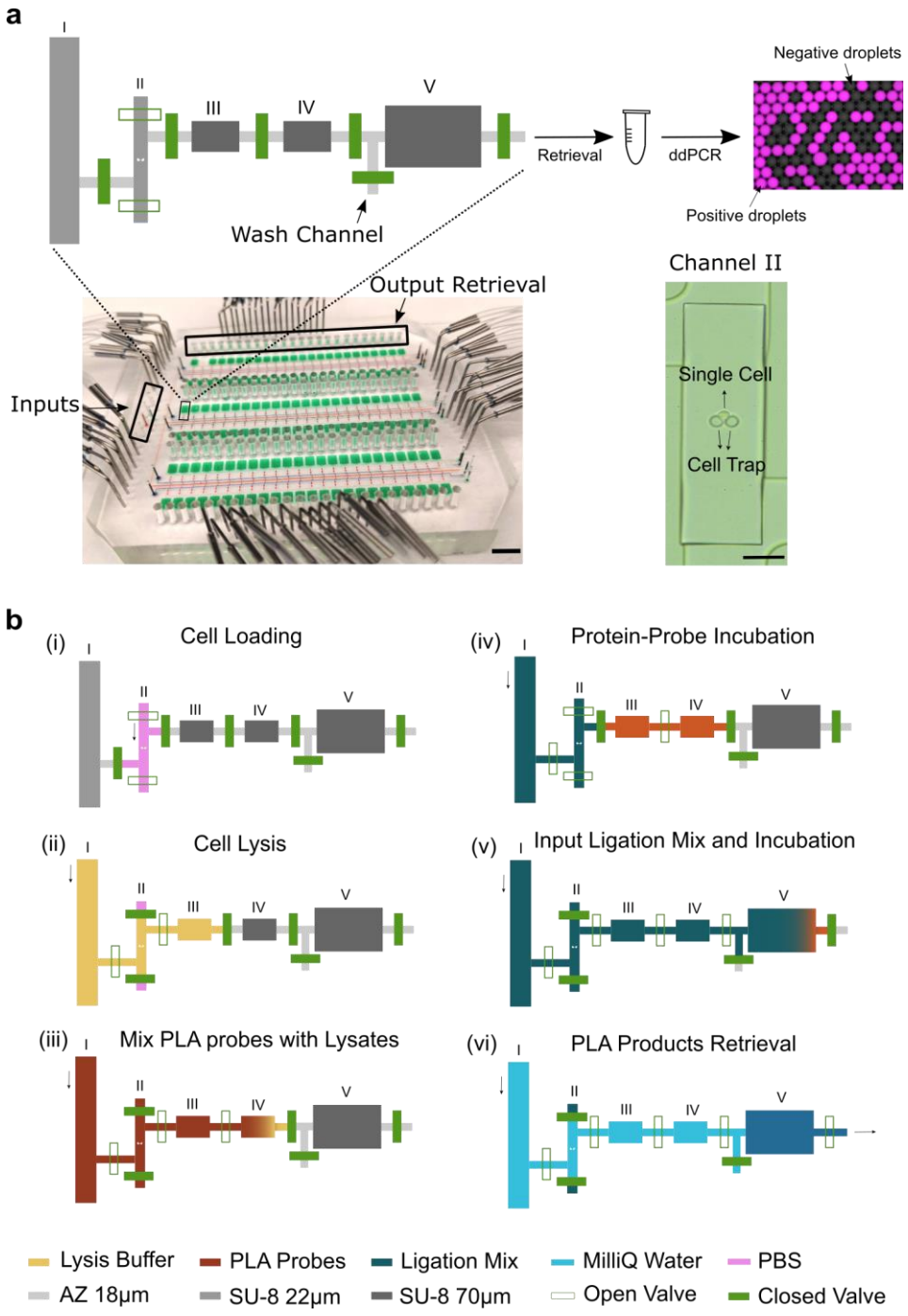

Figure 2-2: Integrated microfluidic device for performing ultrasensitive single-cell protein/messenger RNA (mRNA) measurements. (a) Top: The schematic of one unit of assay chambers is shown; chamber sizes are not to scale. The single cell is trapped in chamber II. The stepwise assay protocol (I–V) results in a digital PCR (dPCR) readout, where counting of positive droplets allows direct quantitation of proteins or mRNA in the sample. Bottom left: The chip image with food dye loaded in different channels, the scale bar is 2 cm. Bottom right: The microscope image of a single human embryonic kidney cell trapped in chamber II, the scale bar is 50 µm. (**b**) Step-by-step workflow of microfluidic-digital proximity ligation assay (µ-dPLA). More information on the chip design and fabrication can be found in the Methods section.

Table 2-2: Comparison of regular (off-chip) and microfluidic (on-chip) dPLA performance.

	dPLA LOD – off-chip (molecules)	dPLA LOD – on-chip (molecules)	dPLA detection efficiency – off-chip (proteins/DNA)	dPLA detection efficiency – on-chip (proteins/DNA)
TNFR1	66,789[1]	2,277	741[1]	33
CD147	55,385[1]	2,795	70[1]	29
CSTB	1,035,991[1]	18,824[1]	5,024[1]	304[1]

[1]Calibration curves are shown in **Fig. 2-1**. Calculation of detection efficiency is explained in the Methods section.

To show general applicability of our technology, we performed single-cell absolute quantification of two protein targets (CD147 and TNFR1) in three different cell types: human embryonic kidney cells (HEK293T), two human lung epithelial cell lines (H1299 and A549) **(Fig. 2-3**). A calibration curve was created along with each single-cell experiment to convert double stranded DNA readouts to absolute protein copy numbers. Single cells were trapped and isolated on chip with an average occupancy of 90%. For measurement of the protein CD147, all of the cells captured by our device were above the LOD. Single cell CD147 counts had an average of 197,785 proteins in HEK293T cells (N = 266), with a range of 3289 ± 1099 to 872,925 ± 13,393 proteins (**Fig. 2-3b, c**; defined as: average ± standard deviation, where standard deviation is calculated as in Eq. (2) in the Methods section, the standard deviation for single cell readings is Poisson errors from ddPCR). The results generally agreed with previous reports for these cells [11]. Single-cell CD147 counts for human lung epithelial cells H1299 (N = 96) and A549 (N = 98) had averages of 327,609 and 179,513 proteins, and ranges of 67,133 ± 2476 to 704,446 ± 7428 and 7417 ± 1737 to 517,949 ± 6717 proteins, respectively (**Fig. 2-3b, c**). About 86% of the single cells (149 out of 174) were above LOD for tumor necrosis factor receptor (TNFR1) protein measurements in our device. TNFR1 is a membrane protein and is the canonical activator of the nuclear factor-κB (NF-κB) pathway, and it has been a notoriously difficult target to quantify due to its low copy number

and membrane-bound nature [20, 21]. We found that single-cell TNFR1 counts had an average of 14,807 proteins in H1299 cells (N = 88), with a range of 6310 ± 1197 to 33,541 ± 1643 proteins (**Fig. 2-3e, f**). Single-cell TNFR1 counts in A549 cell lines (N = 84) had an average of 13,017 proteins, with a range of 2408 ± 927 to 37,099 ± 1382 (**Fig. 2-3e, f**). We note that single cell TNFR1 counts were all below the LOD of dPLA (66,789 proteins, **Table 2-2**). To our knowledge, single-cell TNFR1 protein counts have not been measured before our study. This highlights the advantage of our microfluidic platform for quantification of lowly abundant proteins. Single-cell TNFR1 measurements in HEK293T cells were all below LOD of our device. Together with the finding from a live-cell imaging experiment, where HEK293T cells stimulated with high dose of TNF-α did not result in activation of the NF-κB pathway **(Fig. 2-3g, h and Fig. 2-5**), the results indicate TNFR1 is either not expressed in HEK293T cells or the protein level of TNFR1 in HEK293T cells is too low to respond to TNF-α stimulation.

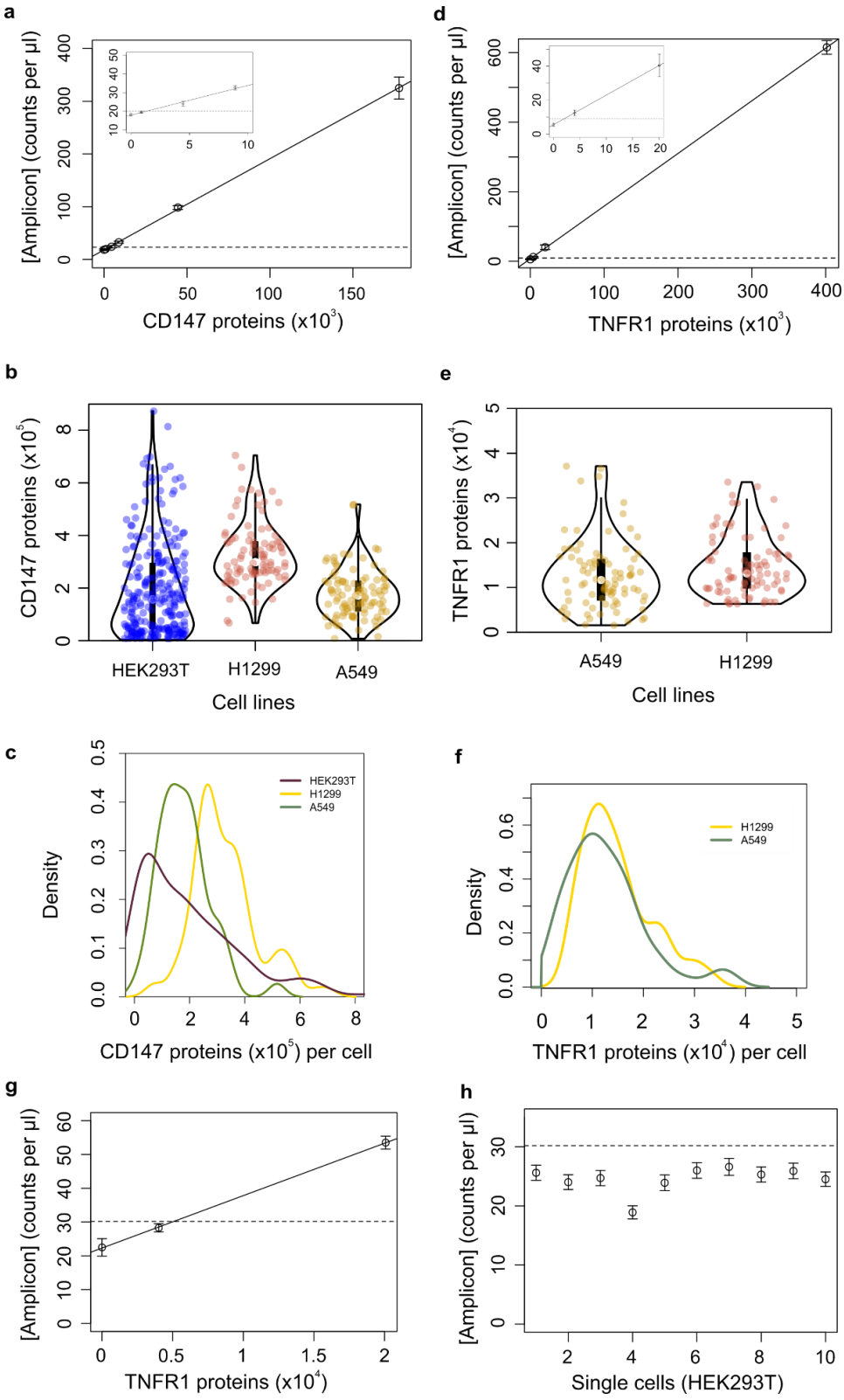

Figure 2-3: Calibration curves and single-cell protein quantification with microfluidic-digital proximity ligation assay (µ-dPLA). (**a**) Calibration curve for CD147 protein (inset: lower concentration

Figure 2-3 (continued): region): $y = 0.00172969x + 18.1043$, $R^2 = 0.9998$, dashed line indicates that limit of detection (LOD) = 2795 protein molecules. (**b**) Violin plots overlaid with scatter plots of single-cell CD147 counts in cell lines: human embryonic kidney cells (HEK293T), human lung epithelial cells (H1299), and human lung epithelial cells (A549). (**c**) Kernel density estimation of single-cell CD147 counts in the three cell lines (because kernel density estimation smooths the distribution, histograms of the single-cell data are plotted and shown in Fig. 2-6). Dark purple line: HEK293T; yellow line: H1299; green line: A549. (**d**) Calibration curve for TNFR1 (inset: lower concentration region): $y = 0.00151339x + 10.4555$, $R^2 = 0.9992$, dashed line indicates that LOD = 2277 protein molecules. (**e**) Violin plots overlaid with scatter plots of single-cell TNFR1 counts in cell lines: H1299 and A549. (**f**) Kernel density estimation of single-cell TNFR1 counts in the two cell lines (histograms of the single-cell data are plotted and shown in Fig. 2-6). Yellow line: H1299; green line: A549. Inside the violin plots, the white dot of the box plot indicates the median of the data, the thick black bar in the center represents the interquartile range (25–75% data range) and the thin black line represents the 95% confidence interval. (**g**) On-chip calibration curve with the number of double-stranded DNA amplicons per microliter plotted against the total protein molecules in the digital PLA reaction: $y = 0.00155149x + 22.3092$, $R^2 = 0.9998$, dashed line indicates the LOD, (**h**) the droplet digital PCR (ddPCR) readings for 10 HEK293T single cells are plotted with respect to LOD calculated from g shown as the dashed line, it shows all the single-cell readings are below LOD. Error bars calculation is mentioned in the Methods section.

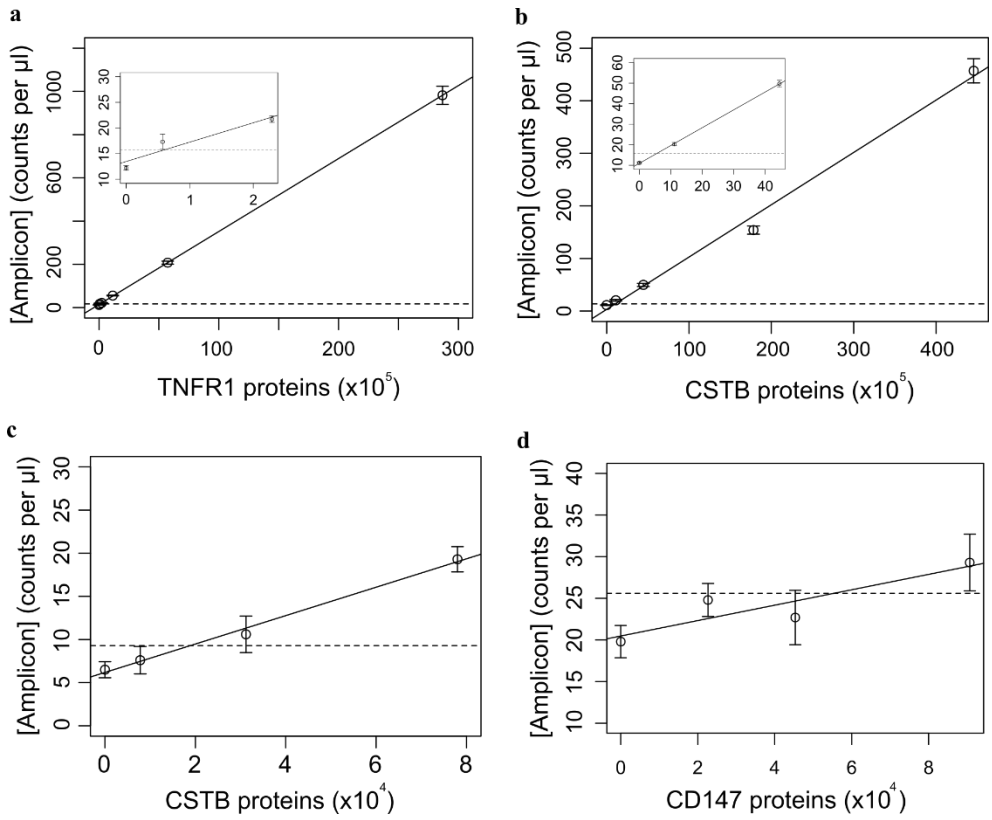

Figure 2-4: Calibration curves for human TNFR1, CSTB and CD147. TNFR1 off-chip (a), CSTB off-chip (b), CSTB on-chip (c), and CD147 off-chip (d) are plotted with number of double-stranded DNA amplicons per µl against the total protein molecules in the digital PLA reaction. Solid lines indicate linear regression for each protein and dotted lines indicate the LOD. (a) TNFR1 off-chip, $y = 3.37\text{e-}5x + 14.5146$, $R^2 = 1$, (b) CSTB off-chip, $y = 9.95\text{e-}6x + 3.14343$, $R^2 = 0.9934$, (c) CSTB on-chip, $y = 0.000164598x + 6.18522$, $R^2 = 0.9929$, (d) CD147 off-chip, $y = 0.92521x + 20.48$, $R^2 = 0.8022$. Error bars calculation is mentioned in the Methods section.

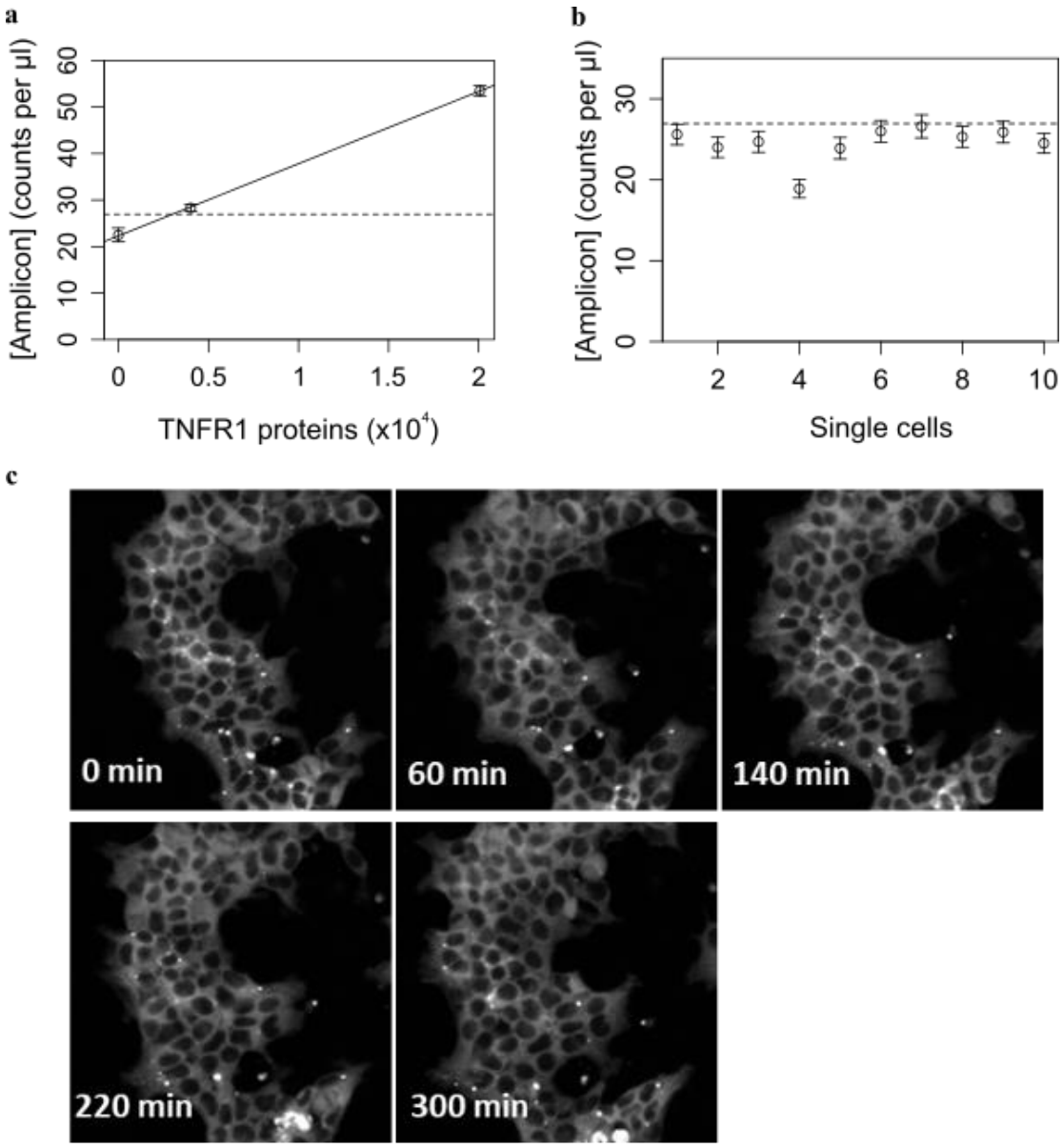

Figure 2-5: HEK293T cells do not respond to TNF- α stimulation. P65-DsRed HEK293T cells are insensitive to TNF-α (8.3 µg ml^{-1}) stimulation, supporting the on-chip digital PLA measurement where no TNF receptor (TNFR1) was detected (calibration curve is shown in **a**; single-cell reads are shown in **b**, where all cells were below LOD). Nuclear translocation of p65-DsRed is expected in stimulated cells, which is absent in the experimental observation (**c**).

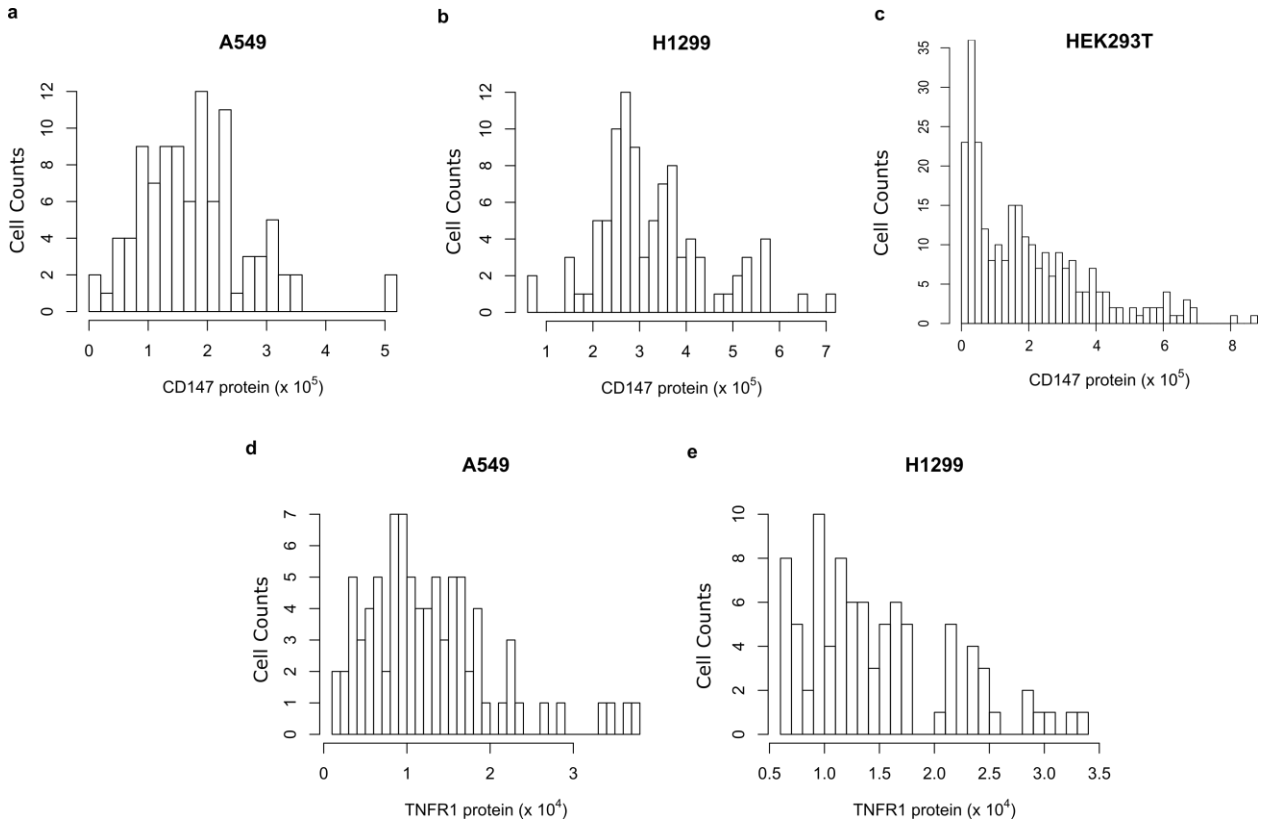

Figure 2-6: Histograms of number of CD147 molecules in (**a**) A549, (**b**) H1299, (**c**) HEK293T single cells; and number of TNFR1 molecules in (**d**) A549 and (**e**) H1299 single cells.

2.2.3 Single-step joint protein/mRNA measurements in single cells

To study and model mRNA/protein expression in single cells, simultaneous measurement of protein and mRNA levels from the same single cell is needed. To achieve such joint measurements in our device, we have combined single-cell mRNA and protein quantification by integrating reverse transcription-ddPCR (RT-ddPCR) with µ-dPLA. Instead of splitting the cell lysates for protein and mRNA measurement separately, we integrated the ddPCR readout for µ-dPLA with one-step RT-ddPCR as a duplex reaction in the same reaction (**Fig. 2-7**). This integrated protocol facilitates mRNA readout without sample bias and interference from protein measurement. TaqMan probe with FAM fluorophore was used for dPLA readout, and VIC fluorophore was used for dPCR mRNA readout. To verify whether the duplex ddPCR reaction creates crosstalk between channels, a series of control experiments were performed, which showed

that VIC and FAM had negligible crosstalk (**Fig. 2-8a, b**). Simultaneous TNFR1 protein (**Fig. 2-7b**) and mRNA (**Fig. 2-7c**) data from the same single cells (H1299) measured with the duplex reaction is shown in Fig. 2-7d (the calibration curve for TNFR1 with the duplex protocol is shown in **Fig. 2-8c**). In short, we achieved simultaneous detection of proteins and mRNA in the same digital PCR droplets, which significantly simplifies the protocols for joint detection of these molecules in single cells.

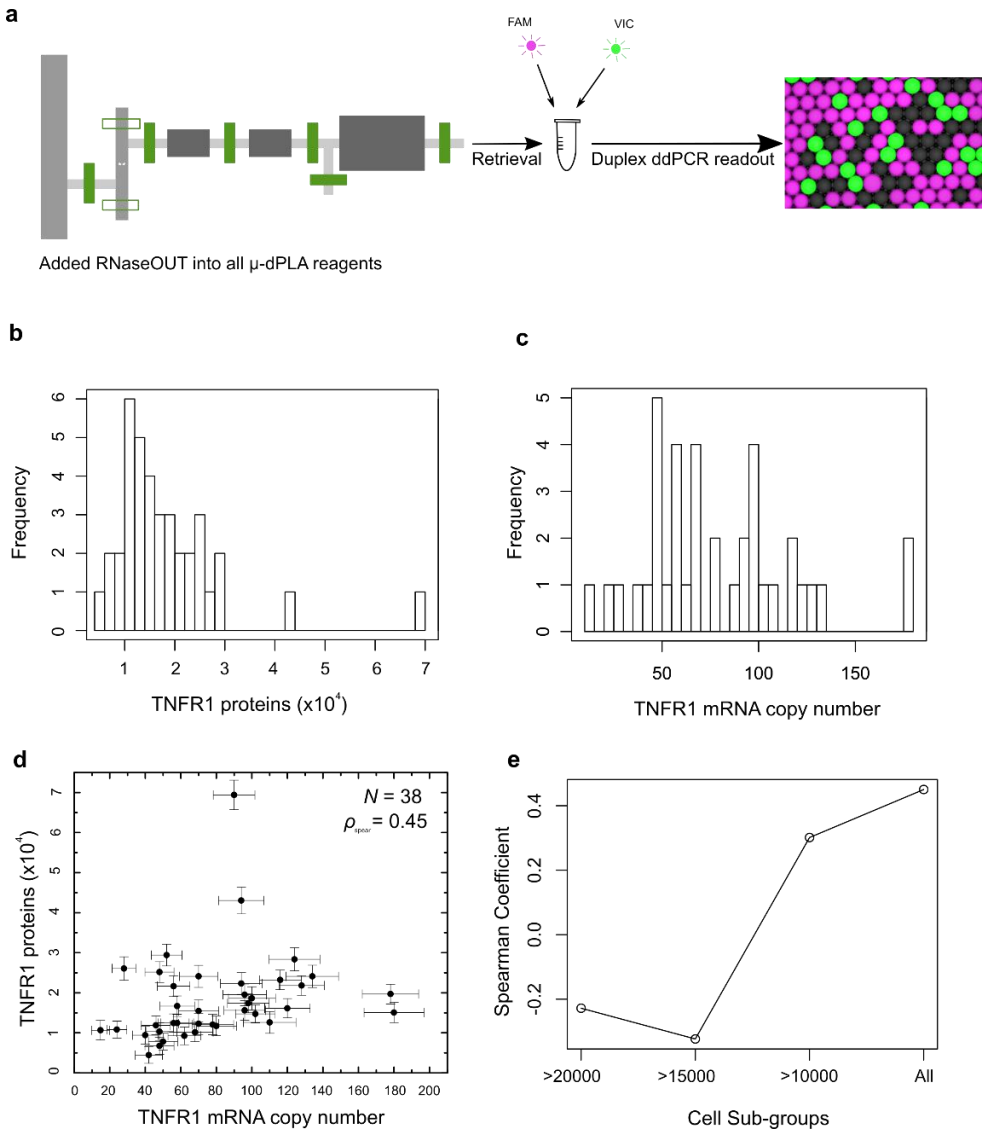

Figure 2-7: Joint digital protein/messenger RNAs (mRNAs) quantification in single cells. (**a**) Workflow of the duplex protein–mRNAs reaction. TaqMan probes with FAM and VIC fluorophore were

Figure 2-7 (continued): used for joint quantification of protein and mRNA, respectively. (**b**) Histogram of TNFR1 protein counts in single human lung epithelial cells (H1299) cells. (**c**) Histogram of TNFR1 mRNA counts in the same single H1299 cells. (**d**) Joint TNFR1 mRNA and protein copy numbers in single cells is shown. See Methods for calculation of error bars. Thirty-eight single cells were measured (N), and the Spearman's correlation coefficient between mRNA and protein counts is 0.45 (ρ). (**e**) The calculated Spearman's coefficient reduces significantly if we exclude the cells that express TNFR1 at very low levels. The X-axis shows the threshold for exclusion, for example, >20,000 indicates that only cells with more than 20,000 TNFR1 protein copies are included in the calculation. Error bars calculation is mentioned in the Methods section.

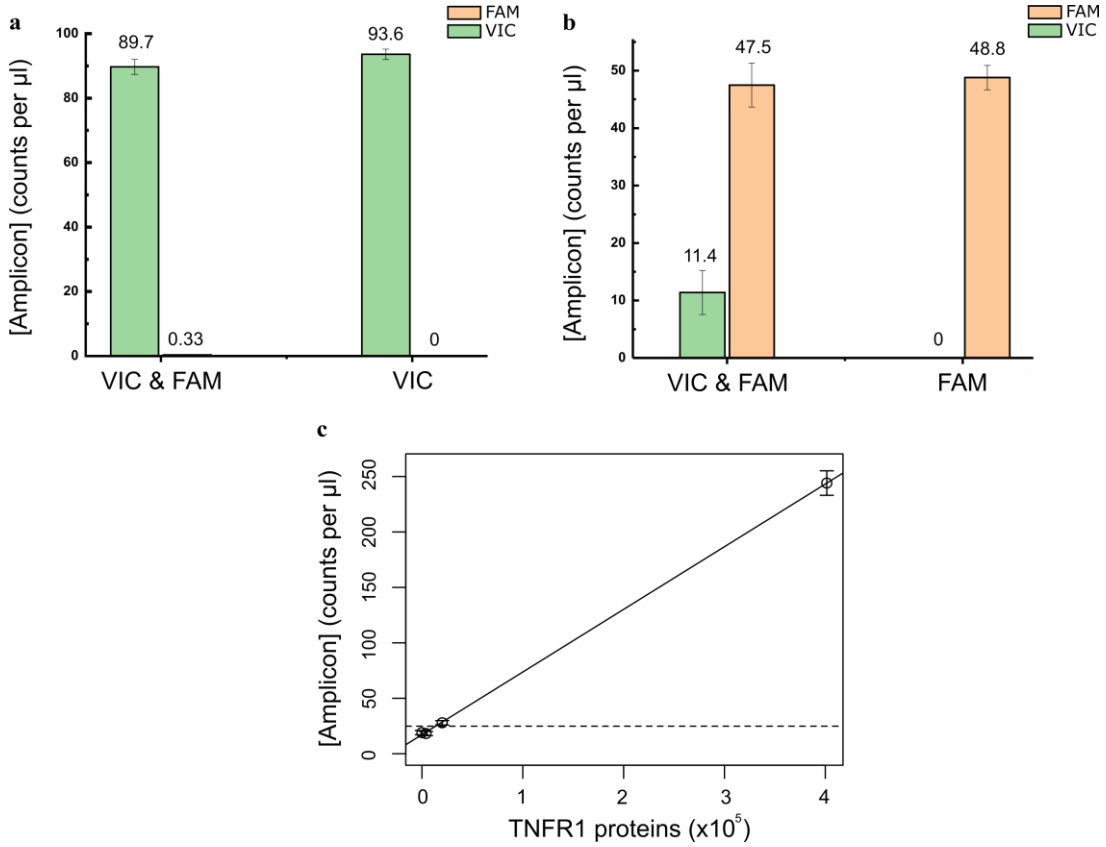

Figure 2-8: Control experiments of protein-mRNA duplex reaction in bulk samples (TNFR1 probes were used), showing there is no crosstalk between FAM and VIC channels. (**a**) TaqMan probe with FAM fluorophore was used for protein quantification, while VIC was used for mRNA quantification. The bar on the left (VIC) shows the triplicate readings of mRNA with only VIC probe, while the ones on the right (VIC & FAM) are with both FAM and VIC probes used in one-step RT-ddPCR protocol. This verifies that with addition of FAM probe, readings in the VIC channel will be approximately the same, and VIC channel will not crosstalk to FAM channel. (**b**) This shows triplicate readings of both DNA products and mRNA from integrated digital PLA & one-step RT-ddPCR protocol, with only FAM probe (right, FAM) against with both FAM and VIC probes used (left, VIC & FAM). This further proves that there is no crosstalk between the two channels, and also digital PLA is compatible with one-step RT-ddPCR protocol. (**c**) On-chip calibration curve for TNFR1 with the duplex reaction protocol, is plotted with number of double-stranded DNA amplicons per µl against the total protein molecules in the digital PLA reaction. Solid lines indicate linear regression for each protein and dotted lines indicate the LOD, $y = 0.000565126x + 17.1769$, $R^2 = 0.9999$. Error bars calculation is mentioned in the Methods section.

2.2.4 µ-dPLA shows increased single-cell mRNA–protein correlation.

mRNA readouts are often used as proxy for protein abundance, but whether transcript levels accurately reflect protein levels is an open question [22]. Our high-sensitivity joint measurements reveal that the single-cell protein and mRNA correlation has a Spearman's coefficient of 0.45 in single mammalian cells for TFNR1 expression (**Fig. 2-7d**), which is significantly higher than many previous reports [11, 14, 23]. We hypothesized that previous reports on low mRNA–protein correlation could be due to limited sensitivity of previous protein measurement techniques, and that sensitive quantification of lowly expressing single cells could increase the mRNA–protein correlation in the population. To test this, we excluded single cells with low protein copy numbers in our measurements and compared the changes in Spearman's correlation coefficient (**Fig. 2-7e**). We found that the mRNA-protein correlation indeed increased as we include more cells with low protein copy numbers. This result shows the importance of high-sensitivity, multi-parameter techniques for accurate understanding of transcription and translation in single cells.

2.2.5 µ-dPLA measures single-cell viral infection heterogeneity.

An advantage of our µ-dPLA system is the capability of integrating live-cell microscopy for high-content single-cell profiling. To demonstrate this capability, we studied human lung epithelial cells (A549) infected with HSV-1 that expresses a yellow fluorescent protein reporter (ICP4-YFP). We infected the cells and monitored the expression of ICP4 protein in single cells by time-lapse measurements in our device. Successfully infected cells showed a nuclear YFP signal, while the un-infected cells (i.e., the cells that received the virus, but infection was not successful) and control cells (the cells that received no virus) did not show any YFP expression (**Fig. 2-9a, b**). After loading the cells on chip and taking fluorescence images, all cells were lysed at the same

time and processed with the µ-dPLA protocol described. We measured HSV-1 proteins with PLA probes using a polyclonal antibody that binds to ~80 viral protein species encoded by the HSV-1 genome. This allowed us to measure the abundance of ICP4 target proteins in individual cells in a single measurement. We observed a significant difference in µ-dPLA protein readout between HSV-1-infected cells (N = 27) and control cells (N = 19) (**Fig. 2-9c**, *p* value of 5.356e-11 calculated from two-sided Welch's t test). ICP4 is a major viral transcription factor, which is expressed immediately upon the entry of the viral genome to the host cell nucleus and prior to viral genome replication [24, 25]. Thus, all YFP-positive cells in our experiments were cells in which the infection proceeded at least to the point of nuclear entry. In these cells, we measured the total viral protein content, composed of all immediate-early, early, and late proteins with µ-dPLA. Our results surprisingly showed that YFP fluorescence (ICP4 expression) is only partially correlated with the total viral protein count (**Fig. 2-9d**, Spearman's correlation = 0.48). This suggests that not all the cells infected by HSV-1 were able to go past the point of immediate-early protein synthesis, and some of these likely represent abortively infected cells [26]. Our method opens the door to a more in-depth analysis of this intriguing sub-population. Furthermore, our single-cell measurements of the viral proteins showed a wide distribution, with the highest expressing cells showing ~5-fold higher viral protein concentration than the lowest expressing cells. This wide distribution is in agreement with previous reports studying viral progeny release from single cells, showing substantial single-cell heterogeneity in virus infection [27, 28].

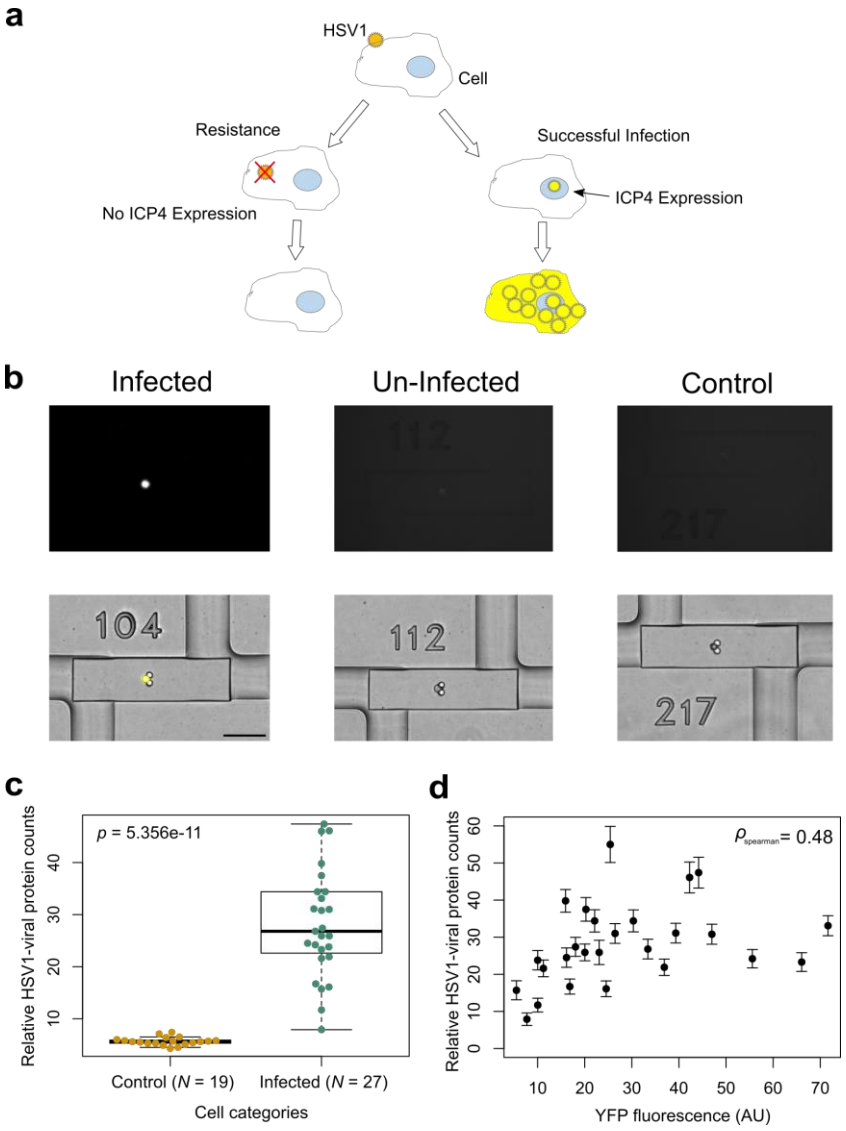

Figure 2-9: Joint live-cell imaging and digital protein analysis of HSV-1 virus infected human lung epithelial cells (A549) cells. (**a**) Schematic of viral infection process: if infection succeeded, viral genome goes into host cell nucleus and ICP4-YFP is expressed and viral proteins are produced; if the cell is resistant to viral infection, ICP4-YFP viral proteins will not be produced. (**b**) Images of ICP4-YFP (top), superimposed bright field (BF) and YFP (bottom) for infected, uninfected, and control single cells (scale bar: 100 µm). (**c**) Comparison of HSV-1 viral proteins counts in control cells (N = 19) and infected cells (N = 27) showed significant differences between two groups. *P* value was calculated using two-sided Welch's t test. The center thick line of the box plot indicates the median of the data, the two hinges of the box represent the first and third quartile, and the whiskers indicate the full data range. (**d**) Correlation of the relative HSV-1 viral protein counts (microfluidic-digital proximity ligation assay (µ-dPLA) readout) with YFP fluorescence level has a Spearman's coefficient of 0.48, which indicates that not all cells infected with HSV-1 went past the immediately-early protein synthesis stage. Error bars calculation is mentioned in the Methods section.

2.2.6 Discussion

Here, we described an automated microfluidic system and associated protocols for ultrasensitive quantification of proteins and mRNA in individual mammalian cells. Our technique, which we called µ-dPLA, significantly improves the sensitivity for single-cell protein profiling: This method enables direct quantification of lowly abundant protein species at a limit of 2,277 protein molecules per cell, within 3 to 4 orders of dynamic range, and detection efficiency of as low as 29 protein molecules. Additional advantages of µ-dPLA compared to existing techniques include rapid processing, and more efficient use of materials and samples. A major advantage of our approach is the combined ability of cell sorting, trapping, and live-cell imaging in time-lapse experiments. With single cells trapped in isolated microfluidic chambers, µ-dPLA spares the use of fluorescence-activated cell sorting, which is bulky and expensive, and is not compatible with the study of rare cells. Additionally, µ-dPLA consumes hundreds of times less reagents and consumables compared to traditional well plate-based assays, which substantially lowers the assay cost.

Population-averaged estimates show that the cellular proteome consists of thousands of very lowly expressed proteins, with copy numbers below 10,000 molecules per cell [29]. These proteins like important regulatory transcription factors constitute an essential part of cellular proteome and hold the key to thorough understanding of cellular functions. However, existing proteomic techniques are limited in their detection limit and sensitivities, which obscures single-cell information and overlooks possible sub-populations. Our method improves the sensitivity of single-cell protein measurements dramatically, allowing the study of many important proteins that are missed with current methods.

The extremely high sensitivity provided by μ-dPLA enabled us to accurately measure the abundance of mRNA and protein molecules in a range of single cells. While earlier reports showed a poor correlation between single-cell mRNA and protein levels (Spearman's coefficient ~ 0.1), we observed a much higher correlation between protein and mRNA levels in single mammalian cells (Spearman's coefficient = 0.45). We have shown that high-sensitivity protein quantification plays a key role in accurately estimating protein–mRNA correlation. As more cells with lowly expressed proteins were included in the calculation, the Spearman's coefficient increased significantly.

Most single-cell analysis techniques provide only single-modality readout, in which researchers need to choose what aspect of the single cells they want to study, such as mRNAs, proteins, or dynamics, while the rest of the cellular information is lost. In contrast, μ-dPLA confers advantages of providing multi-parametric measurement, with integration of duplex one-step RT-ddPCR reaction, live-cell microscopy, and dynamic single-cell stimulation. The integration with live-cell microscopy provides additional information on cell morphology, the expression, and translocation of fluorescent protein reporters, which is useful to decipher the relationships between different protein species or cellular pathways. We took advantage of our combined single-cell proteomic measurements with live-cell imaging and showed that HSV-1 virus infection of lung epithelial cells is highly variable. Our results surprisingly showed that the expression of ICP4, a major transcription factor that regulates hundreds of viral genes, is only partially correlated with the total viral protein counts (Spearman's correlation = 0.48), suggesting that not all the cells infected by HSV-1 were able to go past the point of immediate-early protein synthesis and go through abortive infection. Our combined results highlight the importance of high-sensitivity protein measurements in molecular mechanisms relevant to the study of single cells.

2.3 Methods

2.3.1 Microfluidic chip fabrication

The molds were designed in AutoCAD (Autodesk, USA), and imprinted on photoresist with Heidelberg MLA 150 Direct Write Lithographer (Heidelberg Instruments Mikrotechnik GmbH, Germany). The multi-layer photoresist consists of 18 µm AZ40XT (MicroChem, USA) spun at 1976 × g, 22 µm SU-8 with SU-8-3025 spun at 1372 × g, and 70 µm two-layer SU-8 with SU-8-3025: first layer spun at 1372 × g and second layer spun at 448 × g. The AZ layer went through an overnight reflow process to obtain a rounded shape. The two-layer polydimethylsiloxane (PDMS) device used in this paper is a push-up design, which means the fluidic layer (where reagents are loaded) that is a thick layer bonds on top of the control layer (open/close valves) that is a thin layer. The molds were firstly treated with chlorotrimethylsilane (Cat. No. 92360, SigmaAldrich, USA) for 15 min in a fume hood. This ensured that molds were non-sticky to PDMS, and thus preserved the chip features when peeling PDMS off the molds. To cast the thick layer, 72 g of PDMS was prepared by mixing base and curing agent at 10:1 ratio. The PDMS was mixed and degassed (RTV-615, Momentive Specialty Chemicals, USA) and degassed again after poured onto the molds. To prepare a thin layer, 11 g of PDMS was prepared by mixing base and curing agent at 10:1 ratio. Then, the PDMS was spun on the mold at 448 × g for 1 min. Both layers were then baked at 80 °C for at least 45 min. The thick layers were peeled, punched, aligned, and bonded on top of the thin layer with oxygen plasma (input pressure 860 mTorr) for 18 s (Harrick Plasma, USA). After baking overnight at 80 °C, the fluidic-control-bonded PDMS slab was peeled and punched. The retrieval ports were punched with 1930 µm inner diameter biopsy punch (CR0950765N13R4, Syneoco, USA), and the rest of the holes were punched with 710 µm inner diameter biopsy punch (CR0350255N20R4, Syneoco, USA). Finally,

the device was bonded to a glass slide (127.76 mm × 85.48 mm × 1 mm) with air plasma for 45 s (turn on the plasma at input pressure of 500 mTorr, then briefly refill the chamber with air every 10 s). The bonded device was then baked at 80 °C overnight, before it was ready to use.

2.3.2 Microfluidic chip operation

The schematic of one assay unit is shown in Fig. 2-2. In order to block non-specific binding of reagents to PDMS surface, chips were coated with 10% bovine serum albumin (BSA) (Cat. No. 37525, Thermo Fisher, USA) for 1 h. Chips were then washed with water and ventilated by compressed air before use. The workflow of dPLA for single cells on chip is: (1) Single cells were trapped and isolated in column II. (2) Lysis buffer was introduced from column I to push and lyse the single cell trapped in channel II into chamber III; lysis was completed within 10 min on ice. (3) Columns I and II were washed with PLA probe before they were introduced to mix with the single-cell lysates in chamber III; the mix occupies both chamber III and IV. The chip was then placed at 37 °C for 1 h with the valve between chamber III and IV open, which allows mixing by diffusion. (4) Column I and II were washed with the ligation mix, before it was injected to push the lysate–probe mix (in chamber III and IV) to chamber V. The chip was placed on ice for 30 min for mixing by diffusion, and then placed at 37 °C for 10-min ligation and 65 °C for 5 min to heat inactivate the ligase. (5) The PLA products were flushed out by water and collected for ddPCR readout (**Fig. 2-2b**).

2.3.3 Calibration curve with μ-dPLA

For all targets, a serial dilution of pure protein standards was prepared off-chip, diluted with lysis buffer (0.3× TM buffer). Lysis buffer was prepared by diluting TM buffer and protease inhibitor (Cat. No. K3011010, BioChain, USA) in cell resuspension buffer (Cat No. 4405443, Thermo Fisher, USA) to 0.3× and 1× concentration, respectively. Protein standards were loaded

from the lowest to the highest concentration, with 8-min washing between samples through channel I (**Fig. 2-2**). Then, PLA probes were prepared at a concentration of 500 pM and loaded on the chip, before opening chamber IV (**Fig. 2-2**); channel I was flushed with probes for 2 min, and then channel II was flushed with probes for 15 min. The same flushing protocol also applies to the ligation mix. The rest of the protocol has been described above. After retrieving 9 µl of the final PLA products out of the chip for every unit, they were mixed with 10 µl of 2× ddPCR Supermix for Probes (Cat. no. 1863024, Bio-Rad, USA) and 1 µl of 20× Universal PCR Assay (Cat. No. 4405501, Applied Biosystems, USA). The total 20 µl of sample mix was loaded to DG8 cartridge (Cat. No. 1864008, Bio-Rad, USA) together with 65 µl of Droplet Generation Oil for Probes (Cat. No. 1863005, Bio-Rad, USA). A gasket (Cat. No. 1864007, Bio-Rad, USA) was placed on top of the cartridge holder, and then the cartridge holder was loaded into droplet generator (QX200, Bio-Rad, USA). After that, we transferred the ~40 µl droplets that were generated into a 96- well PCR plate for PCR, which was sealed with an aluminum foil. The PCR program was: 1×, 95 °C for 10 min; 40×, 94 °C for 30 s, followed by 60 °C for 1 min; 1×, 98 °C for 10 min; with the ramp speed as 1.5 °C s^{-1}. The PLA probes for CD147 were conjugated with streptavidin–biotin chemistry, with commercially available oligo kit for PLA (Cat. No. 4448549, Applied Biosystems, USA) and biotinylated antibodies (CD147: BAF972; TNFR1: BAF225, R&D Systems, USA). The PLA probe conjugation protocol is provided in the kit (Cat. No. 4448549, Applied Biosystems, USA). The CSTB PLA probes were commercially available (Cat. No. 4405465, Applied Biosystems, USA).

LOD is defined as the lowest detectable number of protein molecules (calculated as 3 standard deviations of the background reading above either the background reading or the *y*-intercept of the fitted linear curve, whichever is higher). The detection efficiency is defined as the

number of protein molecules converted into one DNA molecule in PLA, and it can be calculated based on the calibration curves as:

$$Detection\ efficiency = \frac{Absolute\ protein\ copy\ numbers}{[Amplicons]\times 20}, \quad (1)$$

where the ddPCR readout ([Amplicon], i.e., amplicons per µl) multiplied by 20 µl gives the total number of amplicons in the reaction.

Error bars indicate ± SD (standard deviation), and SD of the DNA concentration is calculated as [11]:

$$SD = \frac{CI_{max} - CI_{min}}{2\times 1.96} \times \sqrt{N}. \quad (2)$$

Values of CI_{max} and CI_{min} can be retrieved from QuantaSoft result sheet. They are given as TotalConfMax and TotalConfMin for replicate readings, and PoissonConfMax and PoissonConfMin for single readings (such as for single-cell readings). *N* is the number of replicates. We note that the detection limit calculations shown here are based on calibration curves established using purified protein standards as is common practice, and the actual detection limit of cellular proteins might be different.

2.3.4 Single-cell protein quantification with µ-dPLA

HEK293T and A549 (Cat. No. 86012804, Sigma-Aldrich, USA) cells were culture with Dulbecco's modified Eagle's medium (DMEM) media (Cat. No. 11965-092, Life Technologies, USA) supplemented with 10% FBS (Cat. No. F9665, Sigma-Aldrich, USA). H1299 cells were cultured with RPMI-1640 media (Cat. No. 11835-030, Life Technologies, USA) supplemented with 10% FBS. Cells cultured in flasks were dissociated with TrypLE (Cat. No. 12605028, Life Technologies, USA) for 5 min at 37 °C, and then spun down. Cells were then resuspended in sterile

phosphate-buffered saline (PBS), and then filtered with a 40-μm cell strainer (Cat. No. 352340, Corning, USA). The cells were then counted and diluted to a concentration of 2×10^5 per ml for on-chip single-cell loading, with an input pressure of 5 psi. After verifying on the microscope stage that cell loading was complete, the chip was immediately moved on top of ice. Single cells were lysed with 0.3× TM buffer. Lysis was complete within 10 min on ice, and was confirmed visually under a microscope. The rest of the protocol was the same as for calibration curve on chip.

2.3.5 Duplex protein-mRNA digital quantification in single cells

To simultaneously quantify protein and mRNA from the same single cell, the digital PLA part of the protocol was the same, except that we added RNaseOUT (Cat. No. 10777019, Invitrogen, USA) to every reagent mix (the RNaseOUT was diluted 20 times in the final reagent mix). After retrieving 9 μl of PLA products together with mRNA from the chip, the product was mixed with 5 μl of one-step RT-ddPCR supermix, 2 μl of reverse transcriptase, 1 μl of 300 mM dithiothreitol (Cat. No. 1864021, Bio-Rad, USA), 1 μl of 20× Universal PCR assay, 1 μl of 20× TaqMan probe targeting at *TNFR1* gene (TNFR1 TaqMan Gene Expression Assay, Cat. No. Hs01042313_m1, Applied Biosystems, USA), and 1 μl of nuclease-free water. This makes up to a total of 20 μl solution. Droplets were then generated as described above, and transferred to thermal cycler for a RT-PCR program: 1×, 42 °C for 1 h, 1×, 95 °C for 10 min, 40×, 95 °C for 30 s, followed by 60 °C for 1 min, 1×, 98 °C for 10 min; with the ramp speed as 2 °C s^{-1}.

2.3.6 A549 infection with HSV-1 and μ-dPLA measurement

HSV-1 stock was prepared by infecting Vero cells with a multiplicity of infection (MOI) of 0.01 and harvesting 3 days later by three cycles of freezing and thawing, and the stock concertation was determined by plaque assay. A549 cells seeded in a 6-well plate were infected

with 25 µl of the virus in 2 ml fresh medium overnight (MOI 2). The cells were harvested and prepared as described above to be loaded on the chip.

PLA probes for HSV-1 were conjugated in-house with the following chemistry: dibenzylcyclooctyne-PEG4-N-hydroxysuccinimide ester (DBCO-NHS, SigmaAldrich, USA) was dissolved in dimethyl sulfoxide (Sigma-Aldrich, USA) to a concentration of 3 mM. To a microcentrifuge tube was added 9 µl anti-HSV-1 antibody (Cat. No. ab9533, Abcam, USA) and 1 µl of 3 mM DBCO-NHS, mixed thoroughly, and incubated on ice for 60 min. Free DBCO-NHS was purified through a buffer exchange procedure to PBS (Life Technologies, USA) using a 50 µl, 7 K MWCO Zeba Column (Thermo Fisher, USA). Conjugation was verified by measuring absorption at 280 and 309 nm using a NanoDrop ND-1000 spectrophotometer [30].

Probes were made by reacting the DBCO moiety on DBCO-conjugated antibodies with a terminal azide on PLA oligomers. All PLA oligomers were purchased from IDT (PLA oligo A: /5AzideN/CGCATTGCATCGTCTCGTGGGCTCGGCHHHHACHHHHACHHHNGCAGACA TGCGTGATCGCTAAATCGTG; PLA oligo B: /5Phos/TCGTGTCGTGTCTAAAGTCCACAT GCGTACAAAAAAAAAAAAAAAAAAAAAAAAAA/3AzideN/). A solution of PBS with 0.55 mg ml^{-1} DBCO-conjugated antibody and 40 µM PLA oligomer was reacted for 72 h at 4 °C. Unreacted oligomer was removed by using a BSA Removal Kit (Abcam, USA). The final concentration of PLA probes was determined using a NanoOrange Protein Quantitation Kit (Thermo Fisher, USA). Fluorescence was measured at 485/590 nm from a 96-well black glass bottom plate (In Vitro Technologies, Australia) using an infinite M200 Pro (Tecan, Switzerland).

In all, 500 nM PLA probe stock solution was diluted to 2 nM with probe binding buffer (prepared as 0.1% BSA, 0.05% Tween-20, 100 nM goat IgG, 0.1 g l^{-1} sheared salmon sperm DNA in PBS). Then, the ligation mix was prepared by mixing 89.1 µl UltraPure water (Cat. No.

10977023, Invitrogen, USA), 10 µl of 10× ligase buffer (Cat. No. B69, Thermo Scientific, USA), 0.4 µl of 1/500 diluted T4 ligase (Cat. No. EL0011, Thermo Scientific, USA), and 0.5 µl of 1 µM of connector (TTTCGACACGACACGATTTAGGTC). Primers used for ddPCR were: GCA TCGTCTCGTGGGCTC (forward) and TTGTACGCATGTGGACTTTAGACACGACACGA (reverse).

2.3.7 Fluorescence calculation

Matlab function (imfindcircles) for finding circles is used to localize where the single cells are on the bright field images. First, the background was subtracted from the YFP fluorescent images using ImageJ (with function Rolling Ball Background Subtraction), and then the total fluorescence level was calculated by summing all the pixel values on YFP images within the circle dimensions found on bright field images.

2.3.8 qPCR and ddPCR characterization

PLA experiments targeting at TNFR1 were conducted, with side-by-side readout by ddPCR (QX200, Bio-Rad) and qPCR (CFX384, Bio-Rad) for direct comparison of their performances. The same TaqMan probe was used in both PCR readouts. One TNFR1 calibration curve at concentrations: 0.5, 0.1, 0.02, 0.004 and 0.001 ng ml^{-1} was performed to determine the assay LOD with both qPCR and ddPCR. The results suggest that ddPCR produces better readings for PLA than qPCR in terms of both LOD and R^2 (**Fig. 2-1**). The LOD was calculated as background ± 3 SD. We believe the better LOD with ddPCR readout is due to smaller SDs of low concentration and buffer control readings. Then, a series of TNFR1 dilutions with concentration close to the LODs and with small concentration differences (0.012, 0.0115, 0.011, 0.0105, and 0.01 ng ml^{-1}, three replicates for each concentration) were profiled, in order to compare the performances of qPCR and ddPCR at low protein/amplicon concentrations. This is relevant as the

protein/amplicon concentrations are usually low for single-cell experiments and pre-amplification can often introduce bias in the measurements. The results show that in the low concentration range, qPCR sporadically gave no readings. Given that single-cell readings are usually low and we only have one chance to measure a single cell, we conclude that ddPCR is more sensitive and reliable for single-cell measurements. Lastly, we also conducted an experiment targeting at TNFR1 concentrations (1.3, 1.25, 1.2, 1.15, 1.1, and 1.05, 1 ng ml^{-1}, three replicates for each concentration), which are well above LODs and have small concentration differences. This experiment aims to compare the performances of ddPCR and qPCR to resolve small concentration differences. Comparison results were calculated as pairwise *t* test (two-sided Welch's *t* test) between every two concentrations and are shown in **Table 2-1**.

2.3.9 TNF-α stimulation experiment in HEK293T cells

For NF-κB activation in the HEK293T experiment, cells were seeded on an 8-well chamber slide (ibidi GmbH) coated with 0.2 mg ml^{-1} fibronectin (Merck Millipore) and cultured in DMEM high glucose (Thermo Fisher) media supplemented with 10% newborn calf serum (Thermo Fisher) and penicillin–streptomycin (Thermo Fisher). The sample was then mounted on an inverted microscope (Nikon Eclipse Ti) enclosed within a stable cell culture environment at 5% CO_2, 95% humidity, and 37 °C (Life Imaging Services). Time-lapse live-cell images were acquired every 10 min using microscope software (NIS-Elements AR 4.20.01). For NF-κB stimulation experiment, mouse TNF-α (Gibco) was prepared and added to the sample at a final concentration of 8.3 μg ml^{-1}.

2.4 Bibliography

1. Eilken, H.M., S.-I. Nishikawa, and T. Schroeder, *Continuous single-cell imaging of blood generation from haemogenic endothelium.* Nature, 2009. **457**(7231): p. 896-900.

2. Patel, A.P., et al., *Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma.* Science, 2014. **344**(6190): p. 1396-1401.

3. Young, J.W., et al., *Measuring single-cell gene expression dynamics in bacteria using fluorescence time-lapse microscopy.* Nature protocols, 2012. **7**(1): p. 80-88.

4. Zeisel, A., et al., *Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq.* Science, 2015. **347**(6226): p. 1138-1142.

5. Aebersold, R. and M. Mann, *Mass spectrometry-based proteomics.* Nature, 2003. **422**(6928): p. 198-207.

6. Schirle, M., M.-A. Heurtier, and B. Kuster, *Profiling Core Proteomes of Human Cell Lines by One-dimensional PAGE and Liquid Chromatography-Tandem Mass Spectrometry* S.* Molecular & Cellular Proteomics, 2003. **2**(12): p. 1297-1305.

7. Budnik, B., et al., *SCoPE-MS: mass spectrometry of single mammalian cells quantifies proteome heterogeneity during cell differentiation.* Genome biology, 2018. **19**(1): p. 1-12.

8. Bendall, S.C., et al., *Single-cell mass cytometry of differential immune and drug responses across a human hematopoietic continuum.* Science, 2011. **332**(6030): p. 687-696.

9. Stoeckius, M., et al., *Simultaneous epitope and transcriptome measurement in single cells.* Nature methods, 2017. **14**(9): p. 865-868.

10. Peterson, V.M., et al., *Multiplexed quantification of proteins and transcripts in single cells.* Nature biotechnology, 2017. **35**(10): p. 936-939.

11. Albayrak, C., et al., *Digital quantification of proteins and mRNA in single mammalian cells.* Molecular cell, 2016. **61**(6): p. 914-924.

12. Fredriksson, S., et al., *Protein detection using proximity-dependent DNA ligation assays.* Nature biotechnology, 2002. **20**(5): p. 473-477.

13. Gullberg, M., et al., *Cytokine detection by antibody-based proximity ligation.* Proceedings of the National Academy of Sciences, 2004. **101**(22): p. 8420-8424.

14. Darmanis, S., et al., *Simultaneous multiplexed measurement of RNA and proteins in single cells.* Cell reports, 2016. **14**(2): p. 380-389.

15. Whale, A.S., et al., *Comparison of microfluidic digital PCR and conventional quantitative PCR for measuring copy number variation.* Nucleic acids research, 2012. **40**(11): p. e82-e82.

16. Sanders, R., et al., *Evaluation of digital PCR for absolute DNA quantification.* Analytical chemistry, 2011. **83**(17): p. 6474-6484.

17. Kellogg, R.A., et al., *High-throughput microfluidic single-cell analysis pipeline for studies of signaling dynamics.* Nature protocols, 2014. **9**(7): p. 1713-1726.

18. Genshaft, A.S., et al., *Multiplexed, targeted profiling of single-cell proteomes and transcriptomes in a single reaction.* Genome biology, 2016. **17**(1): p. 1-15.

19. Unger, M.A., et al., *Monolithic microfabricated valves and pumps by multilayer soft lithography.* Science, 2000. **288**(5463): p. 113-116.

20. Tay, S., et al., *Single-cell NF-κB dynamics reveal digital activation and analogue information processing.* Nature, 2010. **466**(7303): p. 267.

21. Kellogg, R.A. and S. Tay, *Noise facilitates transcriptional control under dynamic inputs.* Cell, 2015. **160**(3): p. 381-392.

22. Liu, Y., A. Beyer, and R. Aebersold, *On the dependency of cellular protein levels on mRNA abundance.* Cell, 2016. **165**(3): p. 535-550.

23. Taniguchi, Y., et al., *Quantifying E. coli proteome and transcriptome with single-molecule sensitivity in single cells.* science, 2010. **329**(5991): p. 533-538.

24. DeLuca, N.A., A.M. McCARTHY, and P.A. Schaffer, *Isolation and characterization of deletion mutants of herpes simplex virus type 1 in the gene encoding immediate-early regulatory protein ICP4.* Journal of virology, 1985. **56**(2): p. 558-570.

25. DeLuca, N.A. and P.A. Schaffer, *Activities of herpes simplex virus type 1 (HSV-1) ICP4 genes specifying nonsense peptides.* Nucleic acids research, 1987. **15**(11): p. 4491-4511.

26. Ejercito, P.M., E. Kieff, and B. Roizman, *Characterization of herpes simplex virus strains differing in their effects on social behaviour of infected cells.* Journal of General Virology, 1968. **2**(3): p. 357-364.

27. Heldt, F.S., et al., *Single-cell analysis and stochastic modelling unveil large cell-to-cell variability in influenza A virus infection.* Nature communications, 2015. **6**(1): p. 1-12.

28. Russell, A.B., C. Trapnell, and J.D. Bloom, *Extreme heterogeneity of influenza virus infection in single cells.* Elife, 2018. **7**: p. e32303.

29. Schwanhäusser, B., et al., *Global quantification of mammalian gene expression control.* Nature, 2011. **473**(7347): p. 337-342.

30. Gong, H., et al., *Simple method to prepare oligonucleotide-conjugated antibodies and its application in multiplex protein detection in single cells.* Bioconjugate chemistry, 2016. **27**(1): p. 217-225.

Chapter 3

Ultrasensitive multiplexed quantification of protein activities in single cells

3.1 Introduction

Understanding the heterogeneity in a cell population demands sensitive and accurate dissection of single-cell contents. Advances of single-cell technologies in the recent decade enable us to scrutinize molecular species in a quantitative manner [1-4], subcellular dynamics under various stimulus [5-7], lineage tracing and stem cell identification [8, 9]. The findings allow the refining of central dogma of biology with information that were previously obscured by population-average measurement. This is also especially relevant for cancer researches, where baffling complexity of tumor heterogeneity hinders the understanding of intratumoral dynamics and metastasis mechanisms [10, 11].

Single-cell protein analysis methods [12-18] are getting more attention recently, as it's shown that mRNA levels only poorly correlate with protein expression at single-cell level [12, 13, 19, 20]. However, quantification of protein expression in itself proves futile and there is an absolute need to look at the activity state of these biomolecules to shed light on their role in defining a cellular state. Our collaborator has already developed an activity-based protein profiling method (activity-dependent proximity ligation - ADPL), utilizing a family-wide chemical probes coupled to a proximity ligation assay, to look at the activity signature of a single protein or an array of proteins [21, 22]. ADPL has also been successfully applied to molecular phenotyping and proteolytic enzyme activity profiling of bulk primary and metastatic tumor tissues.

Here we report a novel modular platform to look at the activity signatures of a panel of enzymes using a microfluidic-based single-cell activity-based proximity ligation (scADPL) assay. We were successful in demonstrating that the platform can detect activity differences of both cytosolic and membrane proteins at the single-cell level. The precise nanoliter liquid manipulation enabled by microfluidic system offers single-cell measurement signal-to-noise ratio comparable with bulk measurement. Multiplexed scADPL was optimized by systematic condition screening and its specificity and accuracy were validated by inhibitor competition and comparative activity profiling. Multiplexed scADPL was also successful in showing clustering of two different cellular populations based on activity signatures.

3.2 Results and discussion

3.2.1 Molecular properties of family-wide chemical probes

The working mechanism of ADPL involves the use of a small molecule – family-wide chemical probe – that profiles the activities of a group of protein targets belonging to the same family by mimicking the known covalent inhibitors of the targets [23]. The chemical probe is consisted of three parts: warhead, linker and recognition tag (**Fig. 1-4**). The warhead is designed to recognize and bind to the reactive site of the targeting enzymes, and the recognition tag is for downstream pull-down/read-out that is connected to the warhead with the linker. In this work, we are focusing on probing serine hydrolase and cathepsin, for which we use fluorophosphonate (FP-biotin; synthesized in-house) and acyloxymethyl ketone motif - desthiobiotin tag (ActivX) respectively. The structures and working mechanisms of the chemical probes are shown in **Fig. 3-1**. For FP-biotin probe, the fluorophosphonate reacts with the serine site and forms a covalent link (**Fig. 3-1A**); while for acyloxymethyl ketone motif – desthiobiotin probe, it reacts with the cysteine site of the active cathepsin proteases (**Fig. 3-1B**).

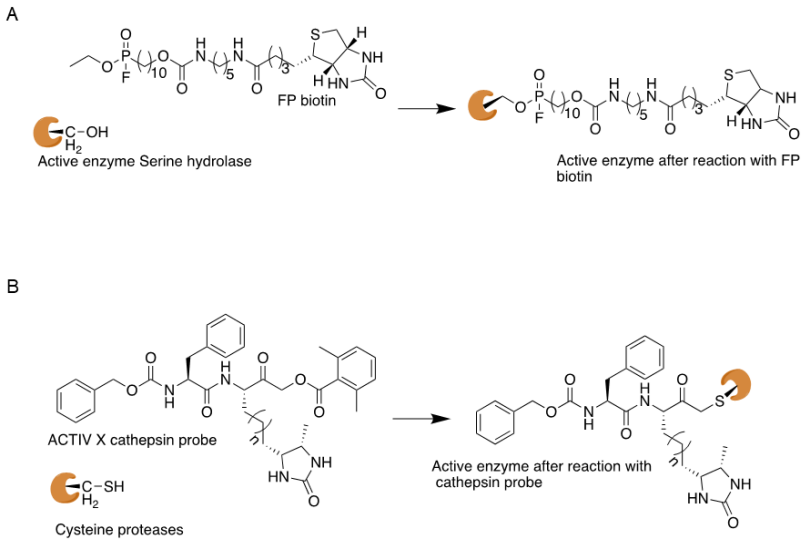

Figure 3-1: **Molecular structures of family-wide chemical probes used in scADPL**. (**A**) Fluorophosphonate (FP) – biotin probe targets at active serine hydrolases. (**B**) ActivX probe targets at active cathepsin proteases.

3.2.2 Rationale and scADPL assay development

As mentioned in the previous section, family-wide chemical probes are widely used to probe the active sites of proteins belonging to the same family (e.g., cathepsin), thus it is desirable to customize the assay such that it is modular with respect to the selection of protein targets we are interested to study. ADPL has therefore been developed to profile activity of a single protein target, by combining family-wide chemical probes with proximity ligation assay (PLA) [21, 22]. For bulk soluble ADPL (sADPL), after regular cell lysis, whole proteome (2 µL) is incubated with a panel of barcoded, protein of interest (POI)-specific antibody-oligonucleotide conjugates and universal streptavidin-oligonucleotide conjugate (**Fig. 3-2A**), which recognizes the biotin recognition tag present on all family-wide probes. Subsequent incubation with complementary, universal splint oligonucleotide allows for ligation of proximal oligonucleotides on antibody (Ab) - and streptavidin (SA) - conjugates (i.e., on the same protein molecule), thereby forming unique

barcoded amplicons that report on the activity of each POI. Finally, ligated amplicons are pre-amplified with universal primers by PCR and detected by real-time quantitative PCR (qPCR) with a barcoded orthogonal primer binding site design [21].

Throughout the entire workflow, an overall ~700-fold dilution was executed in ligation, pre-amplification and qPCR steps, accounting for 99.9% waste of the sample that was not used for the final readout. In addition, the best signal-to-noise point in the linear dynamic range of sADPL requires proteome concentration at 0.1 mg/ml on average [21], while over-dilution could significantly impact on the assay efficiency and limit of detection (LOD), which is especially critical for single cells where protein abundances tend to be low. Considering the mass of a single HeLa cell is ~1.5 ng, single cells should be lysed in ~15 nl solution to achieve the optimal signal-to-noise ratio. To realize the nanoliter liquid manipulation for single cells, we present a microfluidic-based approach, termed single-cell activity dependent proximity ligation (scADPL). The advantages of scADPL over bulk ADPL measurements are: (1) no sample loss; (2) single-cell sensitivity to capture the rare cell populations, which are buried in bulk; (3) less reagent consumption and precise liquid manipulation enabled by automated microfluidic systems; (4) portable with imaging platforms to validate the biological hypotheses by correlating with pre-defined cell types.

Learning from the previous work where the microfluidic platform was used for single-cell protein quantification with digital PLA [20], we adapted our sADPL protocol to achieve optimal performance on the microfluidic chip. Single cells are sorted and isolated in channel II with microfluidic cell trap structure, which is specifically designed to capture mostly single cells; the single cells can then be flushed with different media, stimuli, or buffers as well as imaged for various fluorescent markers before lysis. Then, the single cells are lysed with lysis buffer

introduced from main channel I and lysates are stored in chamber III. Afterwards, Ab/SA-oligo conjugates are also introduced through main channel I, flushed through chamber III and fill up chamber IV; consequently, the lysates and Ab/SA-oligo conjugates mixture occupy both chamber III and IV; we can then proceed to 37 ºC incubation for 1 hour by placing the chip on top of a hotplate with the valve between chamber III and IV open to facilitate mixing by diffusion. Afterwards, we move the chip from hotplate onto ice for another 1 hour incubation at 4 ºC, we have empirically observed that the signal-to-noise ratio has greatly improved with the addition of this step. Then, ligation mixture buffer is introduced through main channel I and pushing the lysates and Ab/SA-oligo conjugates mixture into chamber IV, which is also followed by 30 minutes incubation on ice and 40 ºC incubation for 30 minutes on hotplate. Lastly, the ligation product is flushed out of the chip with UltraPure water, and 9 µl of the product is retrieved with pipetting for downstream PCR readout (**Fig. 3-2A**). More detailed experimental procedures are discussed in Methods.

3.2.3 scADPL assay optimization

Since assay conditions are slightly different between on- and off-chip, we have performed extensive optimization experiments to tune for the optimal conditions, including lysis buffer recipe, Ab/SA-oligo conjugate concentrations, ligation recipe and temperature (**Fig. 3-2D and Fig. 3-3**). To pilot scADPL, we targeted at a panel of 6 enzymes associated with cancer phenotypes in two well-characterized enzyme families, serine hydrolase and cathepsin protease families, using widely available fluorophosphonate-bitoin (FP-Bio) and an acyloxymethylketone-based cathepsin probe, respectively. Using cathepsin L (CTSL) as a model enzyme, we tested a range of Ab-oligo and SA-oligo conjugates concentrations for scADPL in SKOV3 cell lysate, which is an ovarian cancer cell line. To efficiently screen the condition that is compatible with multiple targets, the

optimizations were performed in multiplexed bulk measurement. Whole proteome treated with vehicle and probes were diluted in series and optimized across 5 different temperatures centered at 37 °C for multiple targets. With the increase of the temperature from 30 to 40 °C, the ΔC_T values between vehicle and probes consistently increased for tested targets. Nevertheless, at a higher temperature which may influence the antibody binding, a reversion of difference was observed. Therefore, 40 °C was selected as the optimized ligation temperature. Similarly, the Ab-oligo/SA-oligo concentrations were optimized to be 0.25 nM/4 nM respectively (**Fig. 3-2C and Fig. 3-4**). Furthermore, we looked for a cell lysis strategy that could not only conserve both membrane and cytosolic proteins, but also not interfere with antibody binding with POIs. To effectively compare the dynamic range and signal-to-noise ratio for different conditions, whole proteome treated with vehicle (DMSO) and probes were lysed in different buffers, diluted in a series and processed by multiplexed sADPL (**Fig. 3-3B**). After screening various conditions, particularly regarding the detergent composition, we found that one lysis condition, which we termed MCP buffer, has the best dynamic range and signal-to-noise ratio across all tested enzymes. The MCP buffer was also validated on the microfluidic chip that single cells can be efficiently lysed within few minutes. Lastly, we compared widely used T4 ligase and thermo-stable Ampligase in multiplexed settings (**Fig. 3-5**), Ampligase outperformed T4 in terms of dynamic ranges and working windows.

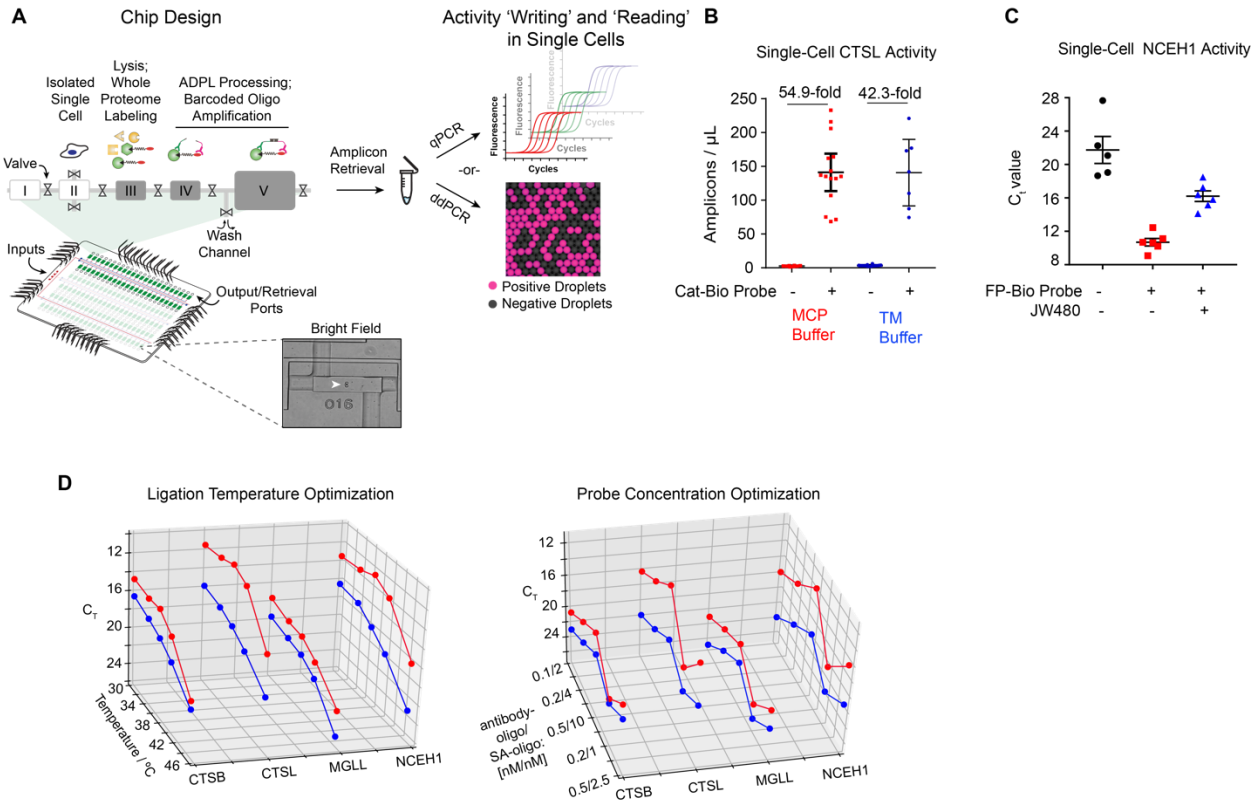

Figure 3-2: Microfluidic chip-based scADPL for activity 'writing' and 'reading' in single cells. (**A**) Chip design and schematic of one unit of assay chambers: input fluid artery (I); single cell trapping (II); cell lysis (III); antibody/streptavidin-oligonucleotide conjugates incubation with whole proteome (IV); activity-dependent proximity ligation (V). Chamber sizes are not to scale. Scale bar for microfluidic chip: 2 cm. Output retrieval containing ligated amplicons is quantified by either multiplexed real-time PCR or droplet digital PCR. scADPL profiles for enzymes CTSL (**B**) and NCEH1 (**C**) in single SKOV3 cells, detected by ddPCR and rtPCR respectively. MCP and TM are two different lysis buffers tested in the experiment, which showed no significant difference in terms of scADPL readings. The recipes for each lysis buffer are mentioned in Methods. NCEH1 signal was abrogated following treatment of live cells with selective inhibitor, JW480. Each data point represents one single cell from 3 technical replicates. (**D**) Assay optimization of ligation temperature and probe concentration by multiplexed bulk measurements. Signal profiles for 4 enzyme targets in SKOV3 cell line were tested at the indicated conditions. C_T values for probes- (red dots) and vehicle- (blue dots) treated proteome were shown in 3D plots. All data points were from technical triplicates in representative biological run. Data represent the mean in **C** and the errors represent standard deviation in **B**.

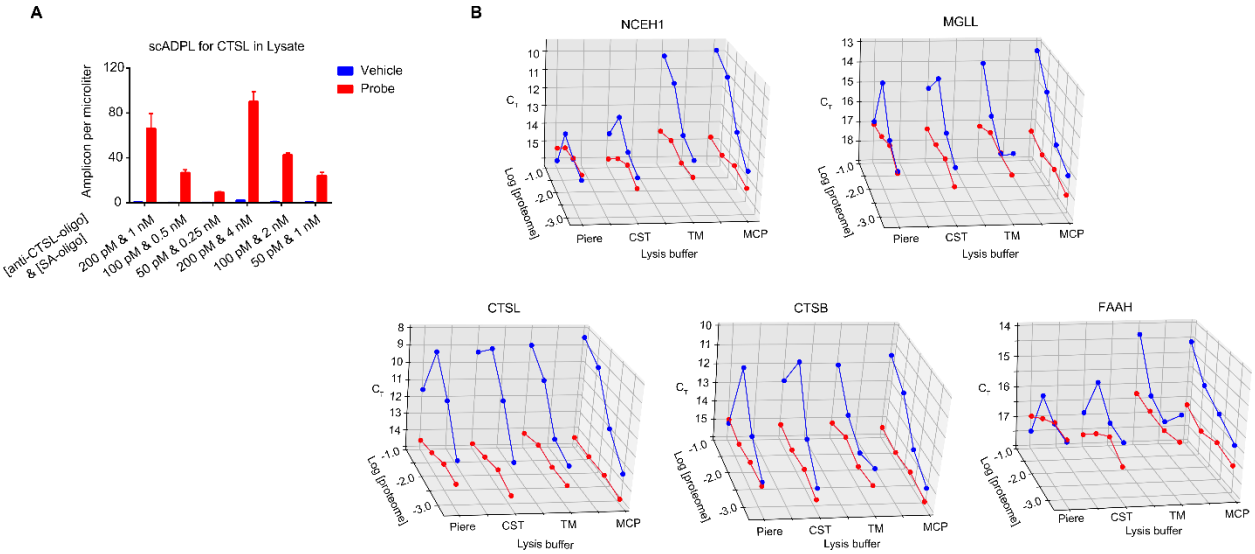

Figure 3-3: scADPL assay condition optimization data. (A) scADPL profiling of active CTSL in SKOV3 cell lysate with the titrated anti-CTSL- / Streptavidin-oligos concentrations. The ligated amplicons were quantified by ddPCR. (B) Cell lysis buffer comparison by multiplexed bulk activity measurements against NCEH1, MGLL, FAAH, CTSB, and CTSL. C_T values for probes- (FP-Bio and cathepsin probe, blue dots) and vehicle- (DMSO, red dots) treated samples at four indicated proteome concentrations were shown in 3-dimensional plot. All data points are from triplicate technical replicates in representative biological experiments. Data represent the mean and the errors represent standard deviation.

3.2.4 scADPL enables ultrasensitive and specific activity quantifications in single cells

With the optimized assay conditions, we then proceeded to examine the sensitivity and specificity of scADPL for enzymes in different enzyme families, including the serine hydrolase enzyme neutral cholesteral ester hydrolase 1 (NCEH1) and cathepsin protease CTSL. The scADPL assay was readout with both ddPCR (CTSL, **Fig. 3-2B**) and qPCR (NCEH1, **Fig. 3-2C**) to show its great compatibility with different readout methods. scADPL profiling in single SKOV3 cells for both targets were successful, with over 50-fold signal-to-noise ratio for CTSL and ΔC_T value of approximately 8 cycles between background and probe treatment for NCEH1 (**Fig. 3-2C**). The specificity of scADPL was verified by small molecular competitive inhibition where live cells were treated with selective inhibitor of NCEH1 (JW480), followed by FP-Bio treatment and scADPL processing. JW480 ablated the scADPL signal for NCEH1 (**Fig. 3-2C**), which demonstrates great assay specificity. To our knowledge, these data were the first time to realize

the activity measurement in single cells and the results suggest that scADPL can be used for both comparative and competitive activity profiling of diverse targets within and between protein families with single cell resolution.

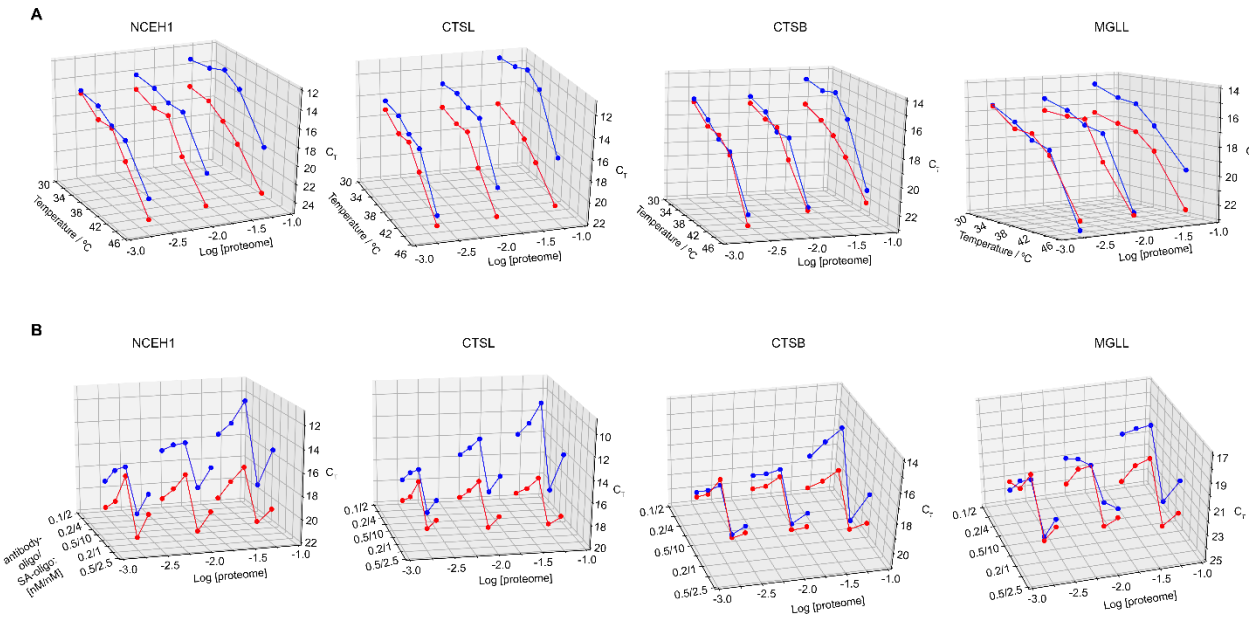

Figure 3-4: scADPL assay optimization. Systematic optimization of ligation temperature (A) and probe concentrations (B) by multiplexed bulk measurements. Signal profiles for 4 enzyme targets in SKOV3 cell line were tested at the indicated conditions. C_T values for probes- (red dots) and vehicle- (blue dots) treated proteome were shown in 3D plots.

3.2.5 Multiplexed scADPL accurately quantifies difference in endogenous enzyme activity

The principle of ADPL design is the conversion of endogenous protein activity into amplified, quantifiable oligonucleotide signals. Our previous study has validated that the barcoding capability of oligonucleotide can be exploited to quantify multiple active enzymes via unique oligonucleotide identifiers. The multiplexed design improved the measurement efficiency, throughput and reduced the sample input requirement. Given the low protein abundances characteristic (thus unfavorable to dilute and split the sample) and one-shot measurement restricted by single cell samples, multiplexed measurement in single cells proves to be more essential than bulk samples to retrieve the information to the maximum extent.

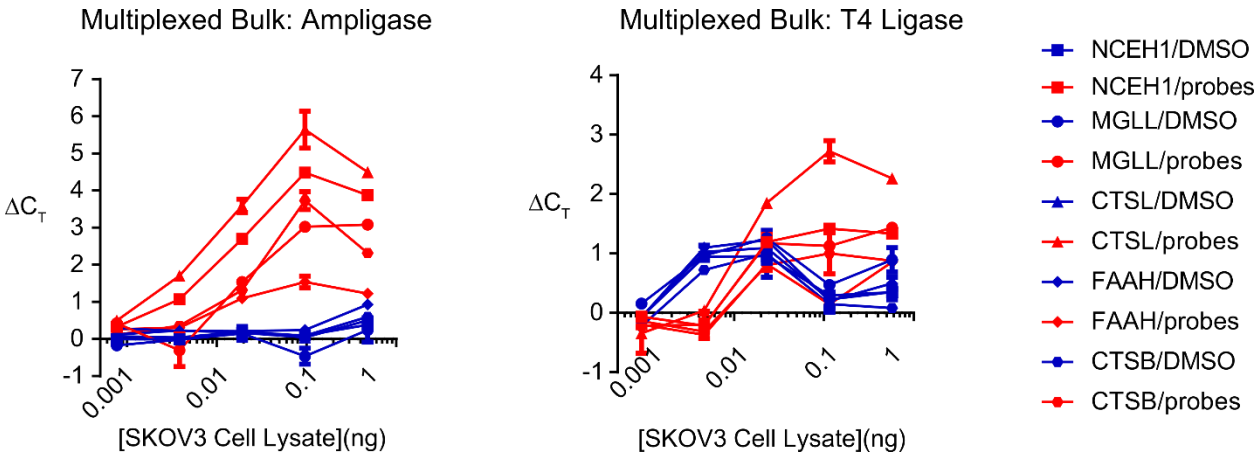

Figure 3-5: scADPL ligase optimization. Optimization of ligase species by side-by-side comparison of Ampligase with T4 ligase measured by multiplexed bulk sADPL. Activity profiles for 5 enzymes across whole proteome dilutions in SKOV3 cancer cell line are shown. ΔC_T, cycle threshold normalized to no lysate (PBS) control. All data points are from 3 technical replicates in representative biological experiments. Data represent the mean and the errors represent standard deviation.

Given the optimized conditions, we performed multiplexed scADPL in single SKOV3 cells to simultaneously analyze diverse enzyme targets, including the serine hydrolase enzymes NCEH1, monoacylglycerol ligase (MGLL), fatty acid amide hydrolase 1 (FAAH), and urokinase (uPA) (**Fig. 3-6A**), as well as the cysteine protease enzymes CTSL and cathespin B (CTSB). Compared against vehicle control, probes treatment generated 2 to 8 PCR cycle differences (**Fig. 3-7A**). The results are consistent with bulk measurement and the activity fold differences were on par with bulk measurement [21]. Furthermore, NCEH1 inhibitor (JW480) only abolished NCEH1 activity in multiplexed scADPL assay, indicating the specificity of the assay. A closer inspection of the activity distribution of individual single cells revealed a unanimous trend for all enzymes with only small extent of staggered activities between cells (**Fig. 3-6B**).

To test whether scADPL could accurately quantify endogenous differences in enzyme activities between phenotypically distinct cell types, we profiled the activity of several enzymes in paired ovarian cancer cell lines SKOV3 and OVCAR3 of high and low aggressiveness respectively, for which activity-based proteomic profiling data have been published in gel-, mass

spec- or imaging-based measurement [21, 22]. The results obtained with scADPL highly agrees with the published results, with all targets being significantly different and SKOV3 demonstrating overall higher activities (lower C_T) across the entire panel (**Fig. 3-7B**).

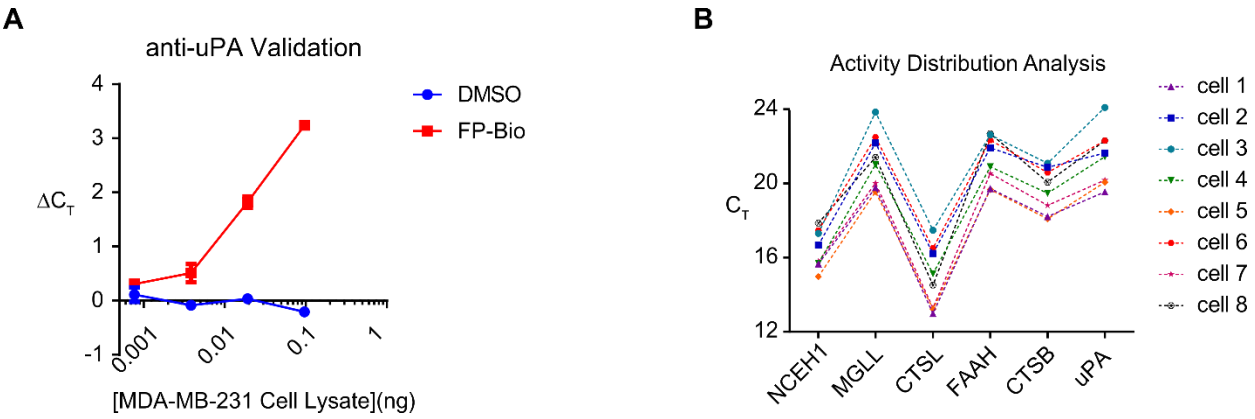

Figure 3-6: Further scADPL probe validation data. (**A**) Urokinase antibody validation by bulk sADPL profiling for urokinase across whole proteome dilutions from MDA-MB-231 cell line. (**B**) Single cell activity distribution analysis measured by scADPL. ΔC_T, cycle threshold normalized to no lysate (PBS) control. All data points are from triplicate technical replicates in representative biological experiments. Data represent the mean and the errors represent standard deviation.

To test the reproducibility of scADPL, two analyses with biological replicates of probes-treated single cells were performed on different microfluidic chips and different days. The correlation result reveals a consistent relative activity trend between two chips, confirming the high reproducibility of scADPL (**Fig. 3-7D, E**). A closer inspection of these measurements however revealed a 2-3 PCR cycles difference of the averaged C_T values for each enzyme. To facilitate direct comparison of single cells analyzed on different chips/days and normalize against unit-to-unit variations on the same chip (i.e., assay efficiency, unit-to-unit volume variations, etc.), we introduced a spike-in control, R-Phycoerythrine (RPE), into our assay. RPE is introduced together with the lysis buffer, delivered to each unit and processed by PLA (instead of ADPL) together with all the scADPL targets. Furthermore, since we are processing about a hundred single cells per chip, to read out 6 enzyme targets plus 1 spike-in control with qPCR, we need to perform at least 2 384-well qPCR plates to cover all the cells from a single run. As we want more

information about cell-to-cell comparisons for the same target than the comparisons among the targets, we performed qPCR to accommodate the same targets for all the single cells on the same qPCR plate, with each qPCR plate also always accommodating the spike-in control reads for all units to account for microfluidic chip- and qPCR plate-variations. We notice that RPE readings from the same chip mostly show C_T variations within 2 cycles, which again demonstrates great consistency and accuracy of our assay and supports the credibility of our data. We therefore use the RPE readouts from each unit to normalize the readings of each enzyme target, and also eliminate the readings that were far off from the average, which generally indicated the assay didn't work properly in that unit (e.g., the unit with RPE reading of C_T value of 30, while the average of all units was 16).

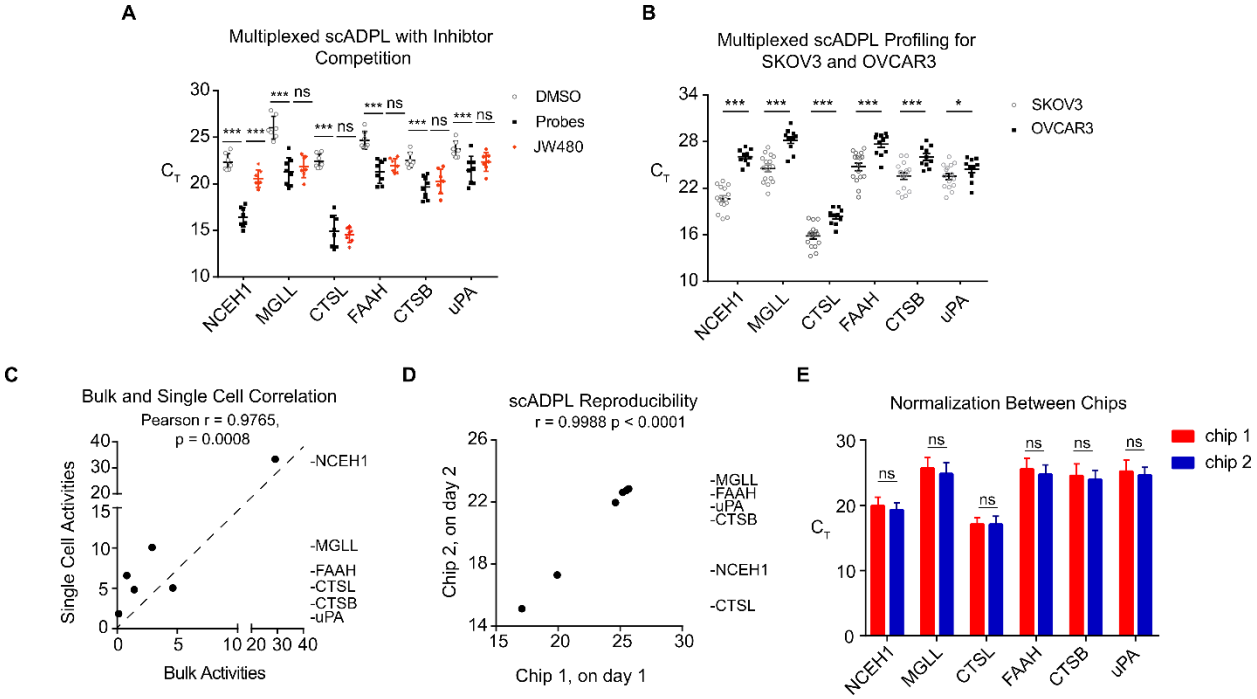

Figure 3-7: Specific, accurate and robust activity measurement in single cells by multiplexed scADPL. (**A**) A representative, multiplexed scADPL profiling experiment detecting 6 active enzymes from single SKOV3 cells. Target-specific engagement of the NCEH1 inhibitor JW480 quantified by multiplexed scADPL. Each data point represents one single cell from 3 technical replicates. (**B**) Comparative scADPL profiling of active enzyme levels between aggressive (SKOV3) and non-aggressive (OVCAR3) ovarian cancer cell lines. (**C**) Plot correlating SKOV3/OVCAR3 activity ratio measured by bulk experiment (*x*-axis) and multiplexed scADPL (*y*-axis). (**D**) Correlation plot of multiplexed scADPL runs for 6 enzyme activity

Figure 3-7 (continued): levels in single (SKOV3/OVCAR3) cells performed on different days and microfluidic chips. Averaged C_T values of multiple single cells (n = 18 for chip 1; n =23 for chip 2) were plotted. (**E**) Normalization to no proteome (PBS) control eliminate the relative day-to-day and chip-to-chip differences. All data points are from triplicate technical replicates from representative biological replicates. Values represent the mean, and error bars represent the SD.

3.2.6 Multiplexed scADPL enables activity profiling in patient-derived cancer organoids

Once we have validated scADPL performance with proof-of-concept experiments, we wanted to apply the pipeline to profile more clinically relevant patient samples to demonstrate its applicability in a more realistic setting. For this purpose, we obtained breast cancer organoids derived from 3 different metastatic sites of a single patient: pleural effusions (left and right lungs), and lymph node metastasis. To profile these samples, the organoids were harvested, dissociated into single cells, and treated with family-wide chemical probes. Then, the treated cells were loaded onto the microfluidic chip for scADPL pipeline. As tumor samples mostly have infiltrating immune cells, we have also performed an extra step of CD45-FITC staining to identify any potential infiltrating immune cells in our samples. Therefore, after loading the sorting the treated cells on chip, we performed fluorescence imaging to differentiate cancer and immune cells. Although we have obtained 3 organoid samples in total, to have more single cells per sample, we only loaded 2 of them on chip – left lung (pleural 1) and lymph node, which showed the largest difference in bulk assay (**Fig. 3-8C**).

The single-cell reads were generally agreeing with the trend we saw in bulk assay – cells from lymph node metastasis showed higher activity profiles than left lung, with all targets demonstrating significant differences (**Fig. 3-8B**). Besides showing the great heterogeneity within and between the samples, the advantages of scADPL over bulk assay would be more significant in fresh clinical samples, where the number of cells tend to be low (e.g., biopsy, insufficient for bulk assays) and the percentage of infiltrating immune cells tend to be much higher (thus could hugely complicate the bulk data).

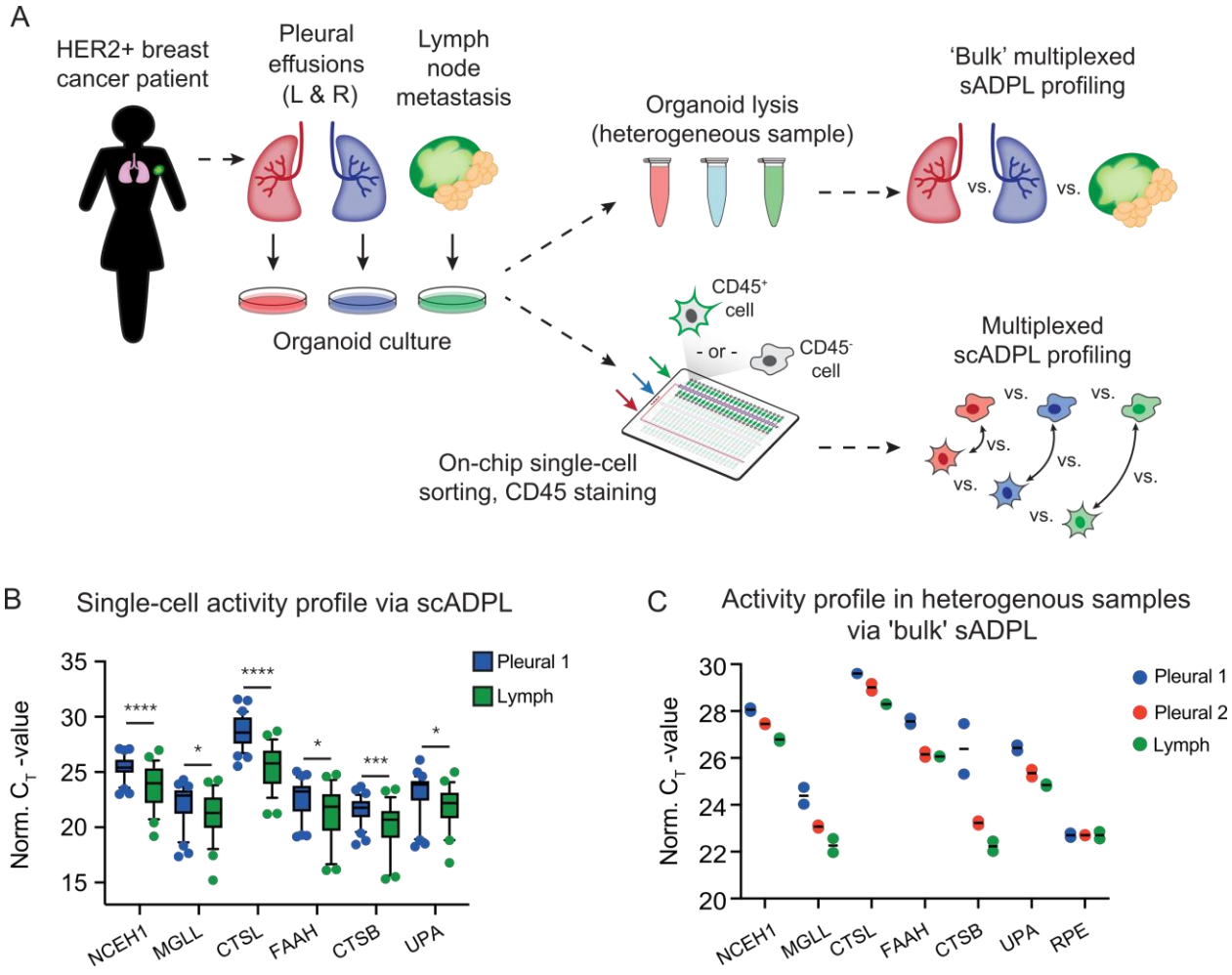

Figure 3-8: Application of scADPL to breast cancer organoid analysis. (**A**) Description of workflow: breast cancer organoids were derived from cells harvested from 3 metastatic sites of a single patient (pleural effusions and lymph node); the organoids were harvested and dissociated into single cells for scADPL profiling. (**B**) Single-cell analysis of breast cancer organoids between pleural 1 (left lung) and lymph node; lymph node shows higher activity than pleural 1 with all targets demonstrating significant differences. (**C**) Bulk analysis of all 3 different organoids, which shows agreeing overall trend with single-cell data in **B**.

3.2.7 Discussion

In this work, we have demonstrated a microfluidic-based platform that enables protein activities profiling in single cells. Instead of profiling the total protein counts, analyzing only the active proteins alone has been shown to be a useful part of cancer prognosis, especially for monitoring metastatic states. On the other hand, single-cell heterogeneity is a dominating phenomenon especially in cancer, where the mixture of highly distinct cell types is complicating

the interpretation of population-averaged analyses. Therefore, profiling active proteins in single cells – our technique termed as scADPL – has the potential to enable much more relevant diagnosis and prognosis tools with unparalleled capabilities.

Taking these factors into consideration, we selected a panel of six cancer metastasis – related targets, and characterized our assay performances in breast and ovarian cancer cell lines in both single-target and multiplexing settings. The assay sensitivity was characterized by comparing single cells treated with vehicle (DMSO) and probes, while the assay specificity was demonstrated with competitive inhibitor for only one of the six targets. Afterwards, the pipeline has been applied to analyze single cells dissociated from patient-derived breast cancer organoids, which could reliably differentiate which metastatic sites the cells originated from.

Although the assay has been demonstrated to work reliably, its application has yet been limited to few biological systems: breast and ovarian cancers. Thus, it is worth being noted that more targets especially from other families (e.g., kinases) are in need for future development to expand the assay applicability. Also, needless to say, it is also imperative to increase the assay multiplexing capability by including more targets in the panel, thus achieving a much wider range of profiling.

On the platform side, the PLA products are currently manually retrieved for individual pre-amplification and qPCR readout for every single cell, which inevitably requires lots of laborious pipetting. Moreover, the excessive pipetting limits the number of single cells that can be processed per run. Thus, it is highly desirable to customize the scADPL pipeline to be compatible with sequencing readout with NGS platforms. Having achieved that, it is also crucial to implement scADPL in an improved microfluidic platform that facilitates easy single-cell barcoding that would

enable pooling all single cells together for sequencing. These improvements would significantly empower the scADPL platform to investigate more obscure clinical questions.

3.3 Methods

3.3.1 Multiplexed soluble activity profiling with sADPL

Cells and patient derived organoids were treated 5 µM of cathepsin family wide probe for 2 hours and 2 µM of family wide serine hydrolase probe FP-Biotin for 30min at 37 °C in RPMI or DMEM supplemented with 10% FBS. Cells are washed with PBS and then lysed in MCP buffer and the protein concentration is determined by a bicinchoninic acid assay (BCA assay). A 3 fold dilution series of the lysates were used for the optimization experiments and for the experiments with organoids, a final concentration of 0.1mg/ml is used. The probes were first treated with PEG-8000 to a final concentration of 5% and then spun down at 4000 rpm for 30 min to remove any potential assay interference. 2 µL of the lysate is mixed with 2 µL of the ADPL probe mixture such that we end up with a final concentration of 200 pM of the enzyme specific antibody oligo conjugate (both for ADPL and PLA) and 4000 pM of the streptavidin oligo conjugates in PBS containing 20 µg/mL poly-A, 2 mM EDTA, 1% BSA, and 0.05% goat IgG. This mixture is then incubated at 37 °C for 1 hour, followed by addition of 116 µL of the ligation mixture 100 nM splint oligonucleotide, 0.05 units of ampligase, 0.3 mM NAD+, 10 mM DTT, and 20 mM Tris-HCl pH 8.3, 50 mM KCl and 1.5 mM $MgCl_2$ and incubation at 40 °C for 30 min. The ligated products were then subjected to pre-amplification PCR by mixing 5 µL of the ligated mixture to 20 µL of the pre-amplification mixture (2.5X buffer, 125 nM primers, 250 µM dNTPs, 9 units Phusion high fidelity DNA polymerase) and amplified for 18 cycles. The pre-amplified products were extracted using 75 µL TE buffer and was subjected to qPCR. 4.5 µL of the extracted pre-amplification mixture was added to 5.5 µL of qPCR mixture (0.9 µM primers and 0.5 µM Taqman

probe in Luna® Universal Probe qPCR Master Mix). Samples were run on a CFX384 Real-Time System (384-well plates). After the PCR the values were either normalized to PBS (for simpler optimization experiments) or normalized to RPE (for more complex experiments).

3.3.2 Bulk lysate and single-cell enzyme activity profiling on chip.

Cells or patient derived organoids were treated with the cathepsin probe, and the FP-biotin as mentioned above. For the singe cell experiment the cells were trypsinized, washed and directly loaded into the chip at a concentration of approximately 2×10^5/mL for single-cell experiments. The cells were then lysed in MCP lysis buffer with or without the addition of the spike-in control RPE on chip. For the bulk lysate experiments on the chip, the cells were lysed in MCP buffer by sonication, following which the protein concentration was measured by BCA assay. The lysates were diluted to the same concentration and then directly loaded into the chip. This is followed by on-chip probe incubation and ligation to give the desired ligation product. The ligated product is flushed out using 9 µL UltraPure water and then pre-amplified using 11 µL of the pre-amplification mixture (2.5X buffer, 125 nM primers, 250 µM dNTPs, 9 units Phusion high fidelity DNA polymerase). The pre-amplified product is then subjected to either qPCR analysis as mentioned in the multiplexed sADPL method, or droplet digital PCR (for single target CTSL readout experiment shown in **Fig. 3-1**).

To perform ddPCR, 9 µL of retrieval sample is mixed with 10 µL ddPCR Supermix for probes (no dUTP) (Cat No. 1863024, Bio-rad), 0.9 µL of 20 µM primers and 0.5 µL of 10 µM TaqMan probe. 20 µL of the mixture is loaded into a droplet generator cartridge (Cat No. 1864008, Bio-rad) "Sample" well, then 65 µL of ddPCR droplet generation oil for probes (Cat No. 1863005, Bio-rad) into the cartridge "Oil" well. After filling all 8 units on the cartridge, place a piece of gasket (Cat. No. 1863009, Bio-rad) on top of the cartridge, and load the assemble into droplet

generator, the process will automatically start once the lid is closed. After droplets are successfully generated, gently transfer the droplets into a 96-well plate (Cat No. 12001925, Bio-rad), covered the well plate with a piece of aluminum foil (Cat No. 1814040, Bio-rad), and sealed at 180 °C for 5 s. Afterwards, transfer the well plate into a thermal cycler set with the following program: 1×, 95 °C for 10 min; 40×, 94 °C for 30 s, followed by 60 °C for 1 min; 1×, 98 °C for 10 min; with the ramp speed set as 1.5 °C s^{-1}.

3.3.3 Microfluidic chip fabrication

The microfluidic chip molds were designed with AutoCAD (Autodesk, USA), and were used as blueprints with Heidelberg MLA 150 Direct Write Lithographer (Heidelberg Instruments Mikrotechnik GmbH, Germany) to expose UV light on 4-inch photoresist-coated silicon wafers to define chip features. The microfluidic chip consists of features of different photoresist heights: 18 µm AZ40XT (MicroChem, USA) spun at 1976 × g, 22 µm SU-8-3025 spun at 1372 × g, and 70 µm coated with 2 layers of SU-8-3025 – first layer spun at 1372 × g and second layer spun at 448 × g. The AZ layer forms a rounded dome-like shape after overnight reflow. The valve-based microfluidic chip consists of three layers: fluidic layer, control layer and glass slide at the bottom. In this paper, we used a "push-up" chip configuration, in which a fluidic layer is thick and bonded on top of a thin control layer that is bonded to a glass slide. The microfluidic chip is made of a polymer material called polydimethylsiloxane (PDMS). Before casting a PDMS device out of the molds, the molds were firstly treated with chlorotrimethylsilane (Cat. No. 92360, SigmaAldrich, USA) for 15 min in a fume hood, which ensured the molds were non-sticky to PDMS, and thus preserve the features on molds for long-term usage. To cast the thick layer, ~72 g of PDMS was prepared by mixing base and curing agent at 10:1 ratio (66 g of base and 6.6 g of curing agent). The mixture was then thoroughly mixed and degassed automatically (RTV-615, Momentive

Specialty Chemicals, USA). The prepared PDMS was then poured over the coated mold and then degassed again. To cast the thin layer, 11 g of PDMS was prepared following the same ratio mentioned before. Then, the PDMS was poured onto coated mold and spun at 448 × g for 1 min. The PDMS-coated thin layer was placed on a level surface for about 15 min to let the PDMS reflow and form an even surface. Both thick and thin layers were incubated at 80 ºC for at least 45 min. The thick layer was then peeled off the fluidic layer mold, punched at the fluidic inlets/outlets, aligned and bonded onto the thin layer with both bonding surfaces treated with oxygen plasma for 18 s (oxygen input pressure 860 mTorr, Harrick Plasma, USA). The aligned assemble was incubated at 80 ºC overnight, then it was peeled off the control layer mold, and was punched again at the control inlets. The retrieval outlets were punched with 1930 µm inner diameter biopsy punch (CR0950765N13R4, Syneoco, USA), and the rest of the inlets/outlets were punched with 710 µm inner diameter biopsy punch (CR0350255N20R4, Syneoco, USA). Lastly, the assemble was bonded to a clean and dry glass slide (127.76 mm × 85.48 mm × 1 mm) with air plasma for 45 s (turn on the air plasma at pressure of 800 mTorr, then briefly refill the chamber with air every 10 s). The final bonded device was then incubated at 80 ºC for at least overnight to ensure tight bonding.

3.3.4 Patient-derived organoid development and preparation

The procurement of biospecimens for the generation of breast cancer (BC) patient-derived organoids (PDO) has been performed according to the approved IRB protocols with a signed patient consent. All patient samples and clinical information related to the samples have been de-identified. Fresh core needle biopsy samples or cells from pleural effusions were collected in excess of all required diagnostic procedures. PDOs were established as previously described [24]. Briefly, tissue obtained from a core biopsy was enzymatically digested for 1h to generate cell

suspension. Cells from pleural effusions were collected by centrifugation at 450 × g for 10 min and washed in Advanced DMEM/F12 containing 10mM HEPES, 1x Glutamax, Penicilin/ Streptomycin (AdvM+). Pelleted cells were mixed with Matrigel (MG), plated into pre-warmed 24 well plates to allow MG to solidify at 37 ºC and supplemented with BC PDO media to grow organoids [24]. To prevent misidentification and cross-contamination, each PDO sample was cultured in a separate, appropriately labeled dish, and the identity of PDOs was confirmed by fingerprinting using The AmpFLSTR Identifiler PCR Amplification Kit. All PDOs were tested for mycoplasma contamination. Once PDOs reached 50-100 µm size, they were collected for subsequent experimental procedures. To dissociate from MG, PDOs were incubated with Dispase at 37 ºC for 20 min, washed twice with AdvM+ and centrifuged at 450 × g for 5 min. To obtain single cell suspension, PDO pellet was incubated with TrypLE at 37 ºC for 5-10 min. Trypsinization was stopped with 2%FBS, cells were washed twice with AdvM+, passed through 40 µm cell strainer and centrifuged at 450 × g for 5 min. Cells were resuspended in 500 µl of BC PDO media and kept on ice before activity profiling on the chip.

3.4 Bibliography

1. Macosko, E.Z., et al., *Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets.* Cell, 2015. **161**(5): p. 1202-1214.
2. Patel, A.P., et al., *Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma.* Science, 2014: p. 1254257.
3. Wu, A.R., et al., *Quantitative assessment of single-cell RNA-sequencing methods.* Nature methods, 2014. **11**(1): p. 41.
4. Islam, S., et al., *Quantitative single-cell RNA-seq with unique molecular identifiers.* Nature methods, 2014. **11**(2): p. 163.
5. Carlo, D.D. and L.P. Lee, *Dynamic single-cell analysis for quantitative biology*. 2006, ACS Publications.
6. Spiller, D.G., et al., *Measurement of single-cell dynamics.* Nature, 2010. **465**(7299): p. 736-745.
7. Tay, S., et al., *Single-cell NF-κB dynamics reveal digital activation and analogue information processing.* Nature, 2010. **466**(7303): p. 267.

8. Kester, L. and A. van Oudenaarden, *Single-cell transcriptomics meets lineage tracing.* Cell Stem Cell, 2018. **23**(2): p. 166-179.

9. Davis, F.M., et al., *Single-cell lineage tracing in the mammary gland reveals stochastic clonal dispersion of stem/progenitor cell progeny.* Nature communications, 2016. **7**(1): p. 1-13.

10. Tang, D.G., *Understanding cancer stem cell heterogeneity and plasticity.* Cell research, 2012. **22**(3): p. 457-472.

11. Marusyk, A., V. Almendro, and K. Polyak, *Intra-tumour heterogeneity: a looking glass for cancer?* Nature Reviews Cancer, 2012. **12**(5): p. 323-334.

12. Albayrak, C., et al., *Digital quantification of proteins and mRNA in single mammalian cells.* Molecular cell, 2016. **61**(6): p. 914-924.

13. Darmanis, S., et al., *Simultaneous multiplexed measurement of RNA and proteins in single cells.* Cell reports, 2016. **14**(2): p. 380-389.

14. Genshaft, A.S., et al., *Multiplexed, targeted profiling of single-cell proteomes and transcriptomes in a single reaction.* Genome biology, 2016. **17**(1): p. 188.

15. Budnik, B., et al., *SCoPE-MS: mass spectrometry of single mammalian cells quantifies proteome heterogeneity during cell differentiation.* Genome biology, 2018. **19**(1): p. 161.

16. Stoeckius, M., et al., *Simultaneous epitope and transcriptome measurement in single cells.* Nature methods, 2017. **14**(9): p. 865.

17. Peterson, V.M., et al., *Multiplexed quantification of proteins and transcripts in single cells.* Nature biotechnology, 2017. **35**(10): p. 936.

18. Rodriques, S.G., et al., *Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution.* Science, 2019. **363**(6434): p. 1463-1467.

19. Taniguchi, Y., et al., *Quantifying E. coli proteome and transcriptome with single-molecule sensitivity in single cells.* science, 2010. **329**(5991): p. 533-538.

20. Lin, J., et al., *Ultra-sensitive digital quantification of proteins and mRNA in single cells.* Nature communications, 2019. **10**(1): p. 1-10.

21. Li, G., et al., *Ultrasensitive, multiplexed chemoproteomic profiling with soluble activity-dependent proximity ligation.* Proceedings of the National Academy of Sciences, 2019. **116**(43): p. 21493-21500.

22. Li, G., et al., *An activity-dependent proximity ligation platform for spatially resolved quantification of active enzymes in single cells.* Nature communications, 2017. **8**(1): p. 1-12.

23. Fang, H., et al., *Recent advances in activity-based probes (ABPs) and affinity-based probes (A f BPs) for profiling of enzymes.* Chemical Science, 2021.

24. Sachs, N., et al., *A Living Biobank of Breast Cancer Organoids Captures Disease Heterogeneity.* Cell, 2018. **172**(1-2): p. 373-386.e10.

Chapter 4

Conclusion and Outlook

In this thesis, I have presented a novel microfluidic platform that enables ultra-sensitive endogenous protein quantifications in single cells. I have transferred a single-cell protein quantification technique, named digital PLA, which was developed in-house onto the microfluidic platform. Since the microfluidic chambers are in the range of nanoliters, they could contain the single-cell proteins in a much higher concentration, comparing to lysing single cells in microliters on well-plates. The new technique, named µ-dPLA, therefore enables quantification of rare protein targets, which may have very low abundances (<10,000 molecules/cell). The microfluidic platform was designed with chambers connected in series, which are precisely tailored with respect to the desired volume ratios, to mimic the bulk assay pipeline. The chambers are separated by sets of microfluidic valves, which can be opened or closed on demand. The 9 µL retrieval was entirely used for final ddPCR readout, which ensures no sample lost therefore maximizing the assay detection limit. µ-dPLA has also been shown to be compatible with simultaneous mRNA-protein quantification from the same single cell without sample splitting. This provides an extra dimension data on the single cells, which is highly desirable. Lastly, µ-dPLA was applied to characterize single-cell heterogeneity in HSV-1 infection. I leveraged the platform compatibility with fluorescent imaging to study the correlation between fluorescence and viral protein counts, which revealed the subpopulation that didn't go past the immediate-early protein synthesis and went through abortive infection (**Chapter 2**).

When I started Ph.D., single-cell protein analysis just started to boom and there were a few interesting works that were dedicated to profile a panel of multiplexed targets in single cell, with less focus on assay sensitivity and accuracy. Leveraging our lab's specialty, which is microfluidics,

I decided to develop an on-chip assay that focuses on improving sensitivity thus enable quantification of rare protein species. I picked targets CD147 because it was benchmarked in our previous digital PLA paper, TNFR1 (TNF-α receptor) because our lab also specializes in NF-κB signaling pathway and it could be useful for our future studies.

With the demonstration of successful TNFR1 quantification in single cells, a natural next step for our lab would be to integrate single-cell TNFR1 quantification and NF-κB activation monitoring to understand the role of TNF-α receptor amount and the chance of NF-κB activation. In the early studies, we have seen stochastic NF-κB activation, and we are still trying to figure out which factors are affecting this behavior the most. To achieve this, we could simply stimulate the cells with p65 fluorescent marker on the microfluidic platform, image the NF-κB activation (nuclear translocation), then quantify TNFR1 with μ-dPLA. The simple experimental setting could potentially provide meaningful insights.

I also collaborated with a chemistry lab to implement an assay with similar pipeline that profiles enzymatic activities onto the same microfluidic platform, which enables ultra-sensitive multiplexed enzyme activity profiling in single cells. The enzyme activity profiling assay, termed as ADPL, probes the activity status of a single POI by combining family-wide chemical probes and PLA. We firstly demonstrated proof-of-concept by performing single target scADPL, which shows clear difference between vehicle- and probe-treated cells with wide working dynamic range. Then the assay has been demonstrated for multiplexing with a panel of 6 enzyme targets in single cells, and the assay specificity was validated with a specific NCEH1 inhibitor named JW480. In a multiplexing readout, JW480 only inhibits NCEH1 reading while the other targets stayed the same. Lastly, the technology was applied to profile single cells dissociated from patient-derived breast cancer organoids (**Chapter 3**).

Currently scADPL is still read out with PCR: after pre-amplification, the sample is diluted and aliquot for separate real-time qPCR readout for each target. Although this method might still allow multiplexing tens of targets, but it requires extensive pipetting for the PCR and ultimately limited to too many multiplexing targets. Therefore, developing scADPL with sequencing readout will be a must if we want to expand the panel of POIs. This way, the number of POIs won't be limited by the sample splitting and number of fluorophores available. On the other hand, it is also imperative to develop new families of chemical probes that are closely relevant to cancer-related mechanisms, such as kinases. With these future improvements, I believe scADPL will be a more powerful tool for cancer diagnosis and prognosis.

APPENDIX

High-throughput transcription dynamics profiling in single cells

Transcription is a highly regulated and dynamic process, and numerous technologies have been developed to profile the cellular transcription status. For instance, RNA sequencing provides genome-wide transcripts counts, however, quantifying only nascent RNA requires special run-on assays [1, 2]. Also, chromatin immunoprecipitation-sequencing (ChIP-seq) identifies and quantifies the segments of DNA where RNA polymerases are bound to [3]. However, it's possible that the polymerases are bound but not performing any actions. Lastly, assay for transposase-accessible chromatin using sequencing (ATAC-seq) [4] probes the segments of the DNA that are unwound from the histones, assuming those segments are undergoing transcription. However, those DNA segments can be opened but not undergoing any process. During the process of transcription or DNA replication, "transcription bubbles" are formed when double-stranded DNAs (dsDNAs) are temporarily opened for access by various enzymes. Therefore, a novel technique termed as kethoxal-assisted single-stranded DNA sequencing (KAS-seq) [5] was designed to probe the "transcription bubbles" specifically, which has been shown to provide more relevant transcription information than other existing techniques.

Currently, KAS-seq has only been demonstrated with population data, which has difficulties resolving transcription and replication information. Therefore, it's imperative to implement a high-throughput platform that enables KAS-seq profiling in thousands of individual cells at the same time. Learning from the single-cell platforms built for other technologies such as single-cell ATAC-seq [6] and single-cell ChIP-seq [7], we decided to implement a similar droplet-based microfluidic platform to achieve single-cell KAS-seq (scKAS-seq). The existing

microfluidic platforms mentioned earlier facilitate the encapsulation and isolation of single cells together with a barcoding agent (either co-encapsulated or injected), such that when the single cells are lysed inside the droplet, the barcoding agent (either barcoding bead or free barcoding oligos) gets hybridized with fragmented DNA, then droplets are gathered, dissolved and barcoded DNAs are pooled for downstream library preparation and sequencing readout.

Learning from these platforms, especially one technology called SNARE-seq [8], we decided to implement scKAS-seq in the following way. Firstly, we treat cells with N_3-kethoxal small molecules simply by adding it into the cell culture media for incubation. The small molecules would penetrate cellular and nuclear membranes, and specifically label guanine in single-stranded DNA (ssDNA). Secondly, we harvest the cells and extract only the nuclei for the next steps. Then, the nuclei are treated with transposases (Tn5), which not only cuts gDNA, but also appends a pair of short DNA segments with 5' overhang on both sides of the fragmented gDNA. It is noted that we only treat fewer than 100,000 nuclei with Tn5 at a time, otherwise Tn5 won't be able to cut gDNA into desired length ranges. Afterwards, the nuclei are harvested again to get single-nuclei barcoded. The barcoding process involves co-encapsulating single nuclei with a barcode bead into the same droplet, lysing nuclei, hybridizing fragmented gDNA and barcodes assisted by a piece of splint oligo (**Fig. A1**), and ligating the barcodes and gDNA fragments. It is then followed by pooling droplets together, breaking droplets, harvesting and purifying barcoded gDNA for downstream library preparation and sequencing.

The barcode design for scKAS-seq mostly derives from Drop-seq barcode design, as it is most widely used and has been validated for various applications. The splint oligo is designed such that it matches partially with the poly-T on the barcode, and also partially with the 5' overhang on the Tn5-fragmented gDNA. This confers significant advantages in capture efficiency as compared

to blunt or AT ligation with fragmented gDNA (for example with fragmentase). Also, learning from the limitations of existing barcoding strategies, such as barcodes are crosslinked onto ceramic beads in Drop-seq, which limits efficiency during PCR amplification step for library preparation, as barcoded gDNAs are still bound on a hard surface that don't have good access to enzymes and primers for efficient reactions. Furthermore, the barcoding strategy in InDrop [9] involves releasing oligos off the beads for better capture and amplification efficiency with UV exposure, which could cleave off the photo-linker that connects the barcode and the bead. However, this could potentially introduce complications as well, as excessive UV could initiate unwanted chemical reactions on DNA and damage the structures. Taking these limitations into consideration, I incorporated a uracil nucleotide at the beginning of the barcode to connect barcode and bead, which can be easily digested by enzymes with close to 100% efficiency without causing any complications to the barcode-gDNA structures. Also, the release-off feature facilitates the KAS pulldown before library preparation, as the small molecule is sensitive to high temperature and denaturing agent, as well as pulling down on a bead is significantly limiting the retrieval efficiency.

To validate whether this barcoding strategy works, I tested it with bulk samples first. To do this, hydrogel beads were synthesis in house following published protocol [10], which accommodate random barcodes (no bead-specific barcodes to save costs for bulk control). Cells were treated with KAS, Tn5 and nuclei were harvested as described above. Then, I took about 20,000 nuclei, spun them down at 600 × *g* for 5 min at 4 °C, removed the supernatant and added lysis buffer (0.2% Sarkosyl, 20 mM EDTA, 0.2 M Tris pH 7.5, 6% Ficoll PM-400; containing 0.5 µM splint oligo: /5Phos/CTGTCTCTTATACACATCTGACGCTGCCGACGAAAAAAAAAA AAAAAAAAAAAAAAAAAAAA), mixed well and incubated on ice for 10 min then 72 °C for 5 min to release Tn5, following by incubation at room temperature with rotation for 30 min for

hybridization of barcode-gDNA-splint to reach steady state. This is followed by extensive washing the beads, to remove the unbound free oligos, thus minimizing background noise. Then, T7 ligation mix (1× T7 ligase buffer, 11,250 units of T7 ligase) was added after removing the supernatant after last wash, incubate at room temperature for 30 min with rotation. After that, beads were washed again and treated with USER mixture (10 units of USER enzyme, 1× USER buffer) at 37 °C for 30 min to release off the barcoded products. Lastly, retrieved the supernatant containing all the barcoded gDNA and cleaned it with DCC kit (Cat No. D4033, Zymo). The final product was then sequenced with Mi-seq (Illumina) and validated that barcode sequences were confirmed to be consistently appended to the beginning of the reads, which proved that our barcoding strategy worked as expected.

To continue this project, a consistent pipeline for in-house hydrogel beads fabrication is needed, as there are no commercially available beads that have release-off features. Alternatively, 10x genomics single-cell multiome kit can also be customized for this application if in-house beads fabrication is not feasible. Next, we need to validate how the quality of scKAS-seq compared to bulk data, before applying the platform for biological applications.

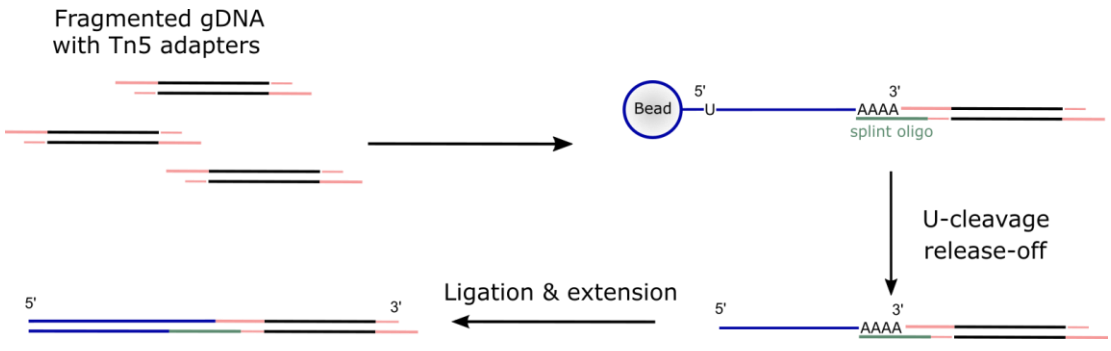

Figure A1: Schematic of barcoding strategy of scKAS-seq. Fragmented gDNA with Tn5 adapters are hybridized to the barcode beads facilitated by the splint oligo. The oligo structure is then released off the bead with USER enzymatic cleavage at the U-nucleotide. The hybridized oligos are then ligated and extended, which is ready for downstream pull-down and library preparation for sequencing readout.

Bibliography

1. Core, L. J., Waterfall, J. J. and Lis, J. T. *Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters*. Science, 2008. **322**(5909): p. 1845–1848.
2. Kwak, H., Fuda, N. J., Core, L. J. and Lis, J. T. *Precise maps of RNA polymerase reveal how promoters direct initiation and pausing*. Science, 2013. **339**(6122), p. 950–953.
3. Park, P.J. *ChIP–seq: advantages and challenges of a maturing technology*. Nature reviews genetics, 2009. **10**(10): p. 669-680.
4. Buenrostro, J.D., et al., *ATAC-seq: a method for assaying chromatin accessibility genome-wide.* Current protocols in molecular biology, 2015. **109**(1): p. 21-29.
5. Wu, T., et al., *Kethoxal-assisted single-stranded DNA sequencing captures global transcription dynamics and enhancer activity in situ*. Nature methods, 2020. **17**(5):p. 515-523.
6. Lareau, C.A., et al., *Droplet-based combinatorial indexing for massive-scale single-cell chromatin accessibility*. Nature biotechnology, 2019. **37**(8): p. 916-924.
7. Grosselin, K., et al., *High-throughput single-cell ChIP-seq identifies heterogeneity of chromatin states in breast cancer.*. Nature genetics, 2019. **51**(6): p. 1060-1066.
8. Chen, S., Lake, B.B. and Zhang, K., *High-throughput sequencing of the transcriptome and chromatin accessibility in the same cell*. Nature biotechnology, 2019. **37**(12): p. 1452-1457.
9. Klein, A.M., et al., *Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells*. Cell, 2015. **161**(5): p. 1187-1201.
10. Zilionis, R., et al., *Single-cell barcoding and sequencing using droplet microfluidics*. Nature protocols, 2017. **12**(1): p. 44-73.