上海交通大學

SHANGHAI JIAO TONG UNIVERSITY

学士学位论文

BACHELOR'S THESIS

论文题目： Highly multiplexed and simultaneous detection of RNA and protein in single cells by mass cytometry

学生姓名： 曹喆燚

学生学号： 3147119008

专　　业： 生物科学（致远荣誉计划）

指导教师： Ruth R. Montgomery，陈广洁

学院（系）： 致远学院

质谱流式细胞术同时检测单细胞中多个 RNA 和蛋白质表达

摘要

2016 年斯坦福大学的研究人员开发了 PLAYR 技术 (Frei AP et al. Nat Method. 2016)，该技术可以使流式细胞仪同时检测单细胞内 RNA 和蛋白质的表达量。本课题开发了新的 PLAYR 技术，可以通过质谱流式细胞术（CyTOF）同时测到单细胞层面的 RNA 和蛋白指标。CyTOF 是一种利用带有金属的抗体检测单细胞内抗原蛋白表达的技术，可以避免流式细胞术的光谱重叠，因此目前的 CyTOF 可以检测单细胞内的 45 个蛋白指标，远多于流式细胞术能检测的数量。我们首先通过细胞表面的抗原表达确定人外周血单核细胞分离出的不同免疫细胞，随后检测不同免疫细胞 *GAPDH* mRNA 的表达量来验证此技术检测 RNA 的能力。此外，用 R848（TLR7/8 的刺激剂）刺激细胞后，利用本课题组开发的技术发现在单核细胞中，IL8,IFNB1,IL1,TNF 基因的 RNA 表达和蛋白表达都上升了，佐证了新技术的可行性。最后，利用新的 PLAYR 技术检测了受到黄热病毒（寨卡病毒和登革热病毒）感染的 PBMC 中一些细胞因子的基因的表达，发现受病毒刺激后单核细胞中 IL8,IFNB1,IL1,TNF,CXCL10 的转录水平和蛋白水平有所提高，表明单核细胞受黄热病毒感染后炎症通路和 I 型 IFN 基因通路激活。上述结果提示新的 PLAYR 技术可以实现用 CyTOF 同时检测被病毒感染的细胞中多个 RNA 和蛋白的表达水平，并可以借此描述不同人群受病毒感染后产生的不同免疫反应，最终可能解释不同病人对病毒有不同敏感程度的免疫学机制。

关键词：RNA，单细胞，质谱流式细胞术，滚环式扩增，黄热病毒

HIGHLY MULTIPLEXED AND SIMULTANEOUS DETECTION OF RNA AND PROTEIN IN SINGLE CELLS BY MASS CYTOMETRY

ABSTRACT

PLAYR (proximity ligation assay of RNA) was developed in 2016 (Frei AP et al. Nat Method. 2016) for transcript quantification by flow cytometry, following the standard antibody staining. We made PLAYR compatible with mass cytometry by time-of-flight (CyTOF). CyTOF is a single-cell analysis technology that uses antibodies labeled with elemental isotopes, hence the lack of interference from spectral overlap between channels shown in flow cytometry. Currently, CyTOF can measure up to 45 parameters simultaneously. Proof of principle of the technology was done by measuring the *GAPDH* mRNA in different immune cell subsets in peripheral blood mononuclear cells (PBMCs), the identities of which were defined on the basis of the epitopes on the cells. Also, the fact that with R848 (Resiquimod, TLR7/8 ligand) stimulation, the expression of *IL8, IFNB1, IL1, TNF* genes were upregulated on both transcription and translation level measured by our PLAYR technology also made our technology convincible. Finally, we applied this technology on the PBMCs infected with flavivirus (Zika virus and Dengue virus) and found the increase of both RNA and proteins of *IL8, IFNB1, IL1, TNF, CXCL10* genes in monocytes infected with Dengue virus, indicating the activation of pro-inflammatory cytokine gene pathway and Type I IFN gene pathway in monocytes infected with flavivirus. This technology enabling the simultaneous and high dimensional investigation of the variations of multiple transcripts and proteins in single cells infected with flavivirus can be applied to characterize the immune

responses to virus infection in various patient cohorts and may help explain the differential susceptibility to flavivirus infection among different people.

Key words: RNA, single cell, Mass Cytometry (CyTOF), rolling circle amplification, flavivirus

Contents

Chapter One Introduction

The ability to measure gene expression in both transcription and translation on single cell level in complex samples, such as virus-infected blood samples, can lead to a system wide understanding of cellular interactions and function. By comparing the various cellular interactions and functions in the blood samples from different patient cohorts, we can investigate the reason of different susceptibility of different patients to the particular kind of virus infection and the different trajectory of the progression of the infections. Eventually, we may be able to predict the various outcomes of different patients infected with a particular kind of virus and make precision medicine possible.

1.1 Existing RNA detection technologies

Single-cell RNA-seq is currently a popular method to study transcript expression in up to several thousand individual cells in a single experiment. ([1] Jaitin, D.A. et al. 2014: Science. 343, 776. [2] Hashimshony, T. et al. 2012: Cell Rep. 2, 666. [3] Islam, S. et al. 2012: Nat. Protoc. 7, 813. [4] Islam, S. et al. 2014: Nat. Methods 11, 163. [5] Ramsköld, D. et al. 2012: Nat. Biotechnol. 30, 777. [6] Sasagawa, Y. et al. 2013: Genome Biol. 14, R31. [7] Shalek, A.K. et al. 2013: Nature 498, 236. [8] Wu, A.R. et al. 2014: Nat. Methods 11, 41. [9] Deng, Q. et al. 2014: Science 343, 193. [10] Picelli, S. et al. 2014: Nat. Protoc. 9, 171. [11] Tang, F. et al. 2011: Nat. Methods 8, S6. [12] Fan, H.C. et al. 2015: Science 347, 1258367.) However, some steps in sample handling (including the separation of live cells before lysis) have been shown to induce significant variations in the transcriptome. ([13] Dvinge, H. et al. 2014: Proc. Natl. Acad. Sci. USA 111, 16802.) Moreover, RNA-seq cannot achieve simultaneous detection of protein epitopes and transcripts, and the usage of this technology is limited by cost, complexity of the protocol and the available sequencing depth and etc. In spite of these limitations, if we

are able to combine the measurement of genome-wide expression and precise quantification of the unique molecular identifiers, we will be able to make single-cell RNA-seq an exceptionally promising technology. ([14] Grün, D. & van Oudenaarden, A. 2015: Cell 163, 799.) As for now, because of the limitation in funding and time, it is reasonable to do a high-dimensional measurement of various particular target transcripts on single cell level in more cells (>10,000) instead of a genome-wide detection of all the RNAs expressed in the samples containing several thousand cells. Moreover, for real clinical usage, single-cell RNA-seq can cost more time and money and should be replaced by other methods if possible. In addition, if we want to know how specific cells respond to cellular signals or diseases, it can sometimes be difficult if they are surrounded by other cell types that don't respond. It's hard to parse out what is real and what could be random signals. If we could look directly at precise populations of cells and their individual responses, we could more accurately target drugs or find vulnerable populations of cells.

Now it is indeed possible to quantify a relatively small number of transcripts while increasing the number of cells to be analyzed. We can use flow cytometry to allow multiple parameters to be measured at a speed of hundreds to thousands of cells per second. Fluorescence in situ hybridization (FISH) protocols can be used to quantify gene expression on flow cytometry platforms. ([15] Bauman, J.G. et al. 1990: J. Microsc. 157, 73. [16] Patterson, B.K. et al. 1993: Science 260, 976. [17] Belloc, F. & Durrieu, F. 1994: Methods Cell Biol. 42, 59. [18] Borzì, R.M. et al. 1996: J. Immunol. Methods 193, 167. [19] Lalli, E. et al. 1992: Anal. Biochem. 207, 298. [20] Just, T. et al. 1998: J. Virol. Methods 73, 163–174.) However, such experiments must have bright FISH signals with good signal-to-noise ratios, since flow cytometry is not able to provide the subcellular imaging resolution necessary to separate individual RNA signals from diffuse background. Recently researchers have used DNA padlock probes

in combination with rolling circle amplification ([21] Larsson, C. et al. 2010: Nat. Methods 7, 395. [22] Weibrecht, I. et al. 2013: Nat. Protoc. 8, 355.) (RCA) and branched DNA technology ([23] Player, A.N. et al. 2001: J. Histochem. Cytochem. 49, 603.) to generate strong hybridization signals. Also, recently the researchers have successfully applied the branched DNA approach to flow cytometry ([24] Porichis, F. et al. 2014: Nat. Commun. 5, 5641.), but only three non-interfering branched DNA amplification systems are available and the spectral overlap of fluorescent reporters will make multiplexing very complicated.

1.2 PLAYR system

In 2016, researchers in Stanford University have developed a technology called PLAYR system. Their PLAYR system addresses these limitations mentioned above by making possible the simultaneous detection of protein epitopes and multiple RNA targets in routine ways in thousands of cells per second. The method is able to preserve input cells in their native states; can detect transcripts without the need for cDNA synthesis; and is compatible with flow cytometry systems. Their technique PLAYR (for Proximity Ligation Assay for RNA) enlists two small DNA probes that recognize adjacent regions of the target mRNA sequence of interest (Figure 1 step 1 and 2). These probes have one region designed to recognize the mRNA, and another side designed to hybridize to two other sequences called backbone and insert (Figure 1 step 3). The backbone and insert are ligated together (Figure 1 step 4) and then copied many times using a special enzyme called phi29 that starts rolling circle amplification using a single strand of circular DNA (Figure 1 step 5). This allows them to create many copies of the backbone and insert for every mRNA they are looking for. Finally, they add a fluorescent DNA probe that recognizes the amplified sequence (Figure 1 step 6). The cells that express the mRNA of interest will now shine a specific color because of the

DNA probe. The probes they use to recognize the mRNA shown in step 2 are very specific. This means there is low background, so it can be ensured that the signal seen is real. Different pairs of probes can also be used to target the same mRNA, ensuring the signal is robust and precise. The PLAYR technique can be done in conjunction with the cell surface markers, which allows you to confirm mRNA expression and quantity in specific cell types ([25] Frei, A.P. et al. 2016: Nat Methods. 13(3), 269.).

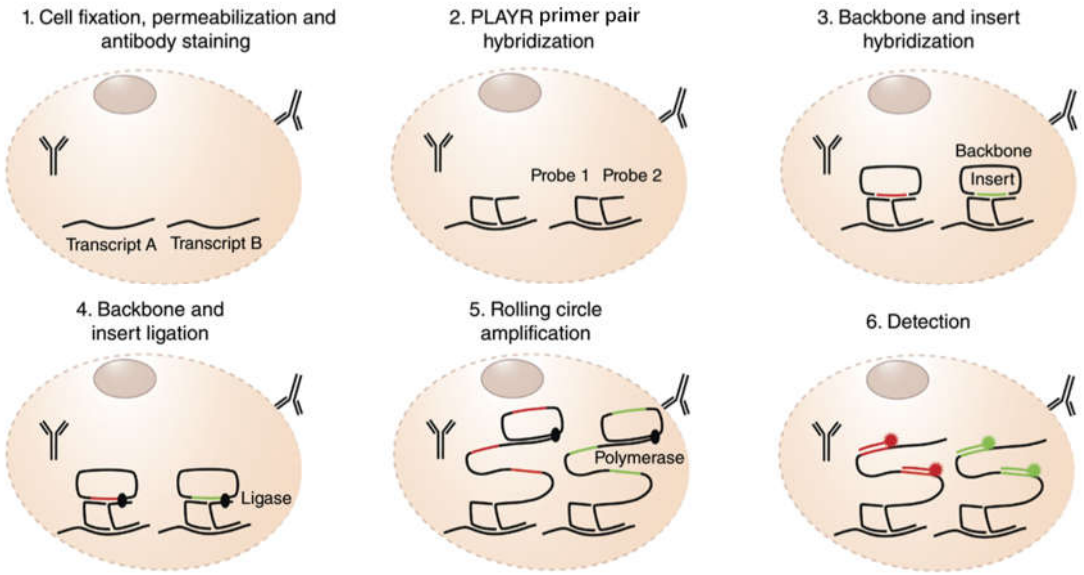


Figure 1 PLAYR enables the simultaneous quantification of specific transcripts and proteins in single cells

The main steps of the original PLAYR protocol. In the cells, in which the target proteins have been stained with antibodies that are ligated with fluorophores, target transcripts are hybridized with the designed primer pairs. Only when both of the primer pairs have been successfully bound with the target transcripts can they be then hybridized with the Insert/Backbone oligos, which will later on be ligated with each other (the sequence of the Insert oligo and the primer pairs are reverse complimentary). After that, the primer pairs will undergo rolling circle amplification to provide multiple binding targets for the detection probes that have been ligated

with fluorophores. After the whole PLAYR procedures having been conducted, the target protein and transcripts in single cells can be detected by flow cytometry.

By using this technique, the researchers are able to distinguish different cell types based on their cell surface proteins, and then confirm their findings by looking at each cell type's gene expression. In their results each transcript shows robust cell-type specificity -- meaning each mRNA or protein appears to only be expressed in the subset of cells that should express them (Figure 2). This data act as a good control, showing that the mRNA labeling works just as well as the more established protein protocols. On the protein level, single-cell measurements have prognostic and diagnostic value in multiple clinical settings. ([26] Irish, J.M. et al. 2004: Cell 118, 217. [27]. Gaudillière, B. et al. 2014: Sci. Transl. Med. 6, 255ra131. [28] Jaye, D.L. et al. 2012: J. Immunol. 188, 4715. [29] Kaleem, Z. et al. 2003: Arch. Pathol. Lab. Med. 127, 42.) PLAYR extends such analysis to include transcript measurements and can supplement the use of antibodies, especially where exon-specific expression is concerned and no relevant antibody reagents exist. Immediate measurement of mRNA can overcome issues introduced with ex vivo processing of live cells, and experimental artifacts are minimized, as PLAYR detects RNA molecules directly without the need for cDNA synthesis

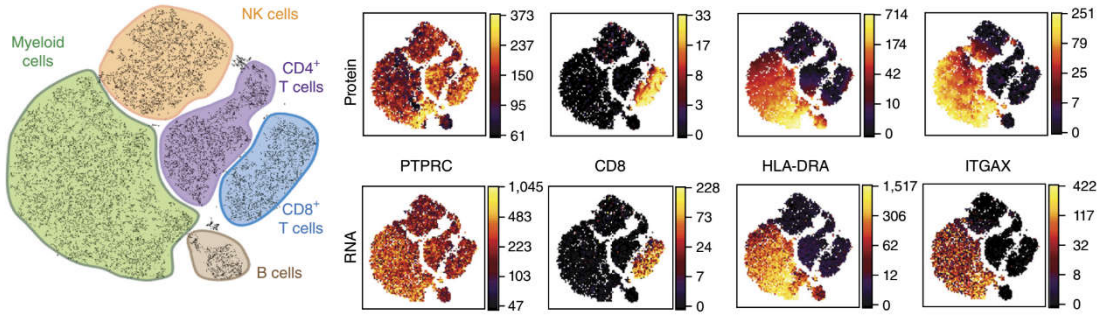


Figure 2 Highly multiplexed measurement of transcripts in cell types defined by protein epitopes. ([25] Frei, A.P. et al. 2016: Nat Methods. 13(3), 269.)

The result from the experiment applying the original PLAYR protocol on human PBMCs. With the original PLAYR procedures being able to simultaneously detect RNA and proteins in the cells by flow cytometry, the whole human PBMCs can be clustered into different cell populations according to the detected epitopes on the cells and the RNA and proteins inside the single cells can also be detected. In this way, we can detect the gene regulation in the specific cell populations on both RNA and protein levels.

1.3 Mass cytometry (CyTOF)

Mass cytometry is similar to the more well-known flow cytometry which involves using fluorescent antibodies that recognize cell surface molecules to distinguish cells types (Figure 3). For example, if you have a homogenous mixture of blood cells like you do in your circulating peripheral blood, like B-cells, t-cells, myeloid cells, etc., you can use fluorescent antibodies that tag surface molecules like B220 on B-cells, or Mac-1 on myeloid cells. In this way, we know all of the cells expressing, say, a green B220 are B-cells, and all of the myeloid cells express a red Mac-1. By using flow cytometry, you can analyze the fluorescence of each single cell to categorize and sort them. Mass cytometry is different in that it doesn't use fluorescent markers, but uses metals as markers. Different metals react differently to electric fields, which the machine can measure to determine which marker is being used. The concept is very similar to flow cytometry, although mass cytometry allows researchers to visualize about two times as many markers as flow cytometry because it doesn't involve distinguishing between similar colors. These machines are capable of analyzing over a thousand cells per second. It is an extremely useful way to analyze bulk populations of cells based on their cell surface markers, allowing you to parse out different cell types. ([30] Bendall, S.C. et al. 2011: Science 332, 687.)

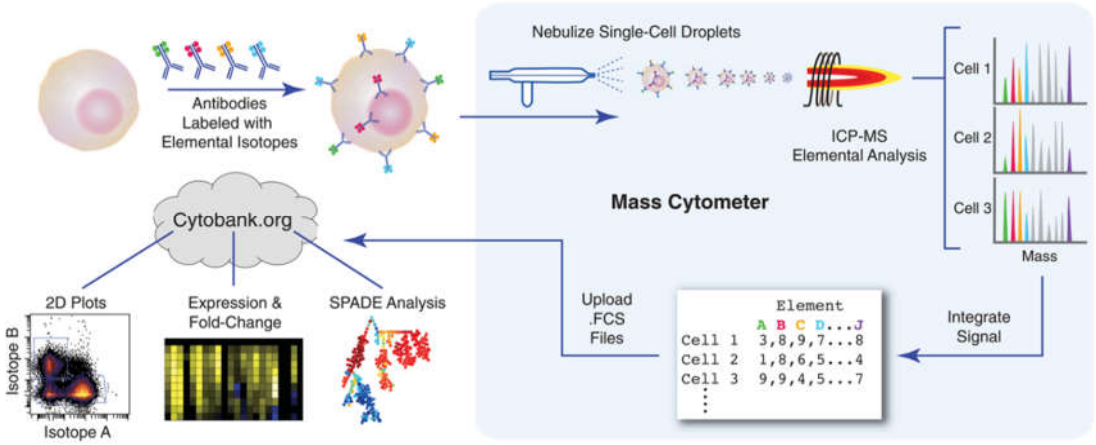


Figure 3 Mass cytometry profiling of immune cell response patterns ([30] Bendall, S.C. et al. 2011: Science 332, 687.)

Workflow summary of mass cytometry analysis. Cells are stained with epitope-specific antibodies conjugated to transition element isotope reporters, each with a different mass. Cells are nebulized into single-cell droplets, and an elemental mass spectrum is acquired for each. The integrated elemental reporter signals for each cell can then be analyzed by using traditional flow cytometry methods as well as more advanced approaches such as heat maps of induced phosphorylation and tree plots.

1.4 PLAYR system compatible with mass cytometry

Building off of this technique, we developed a way to analyze cells using mass cytometry according to their mRNA transcripts – a PLAYR system that is compatible with mass cytometry. We expect that PLAYR will lead to a better understanding of stochastic processes in gene expression and allow for deeper insights into complex cell populations. Using PLAYR with mass cytometry paints a precise picture of transcription at a single cell level. Detecting RNA levels is a more accurate way of measuring cellular responses as they are the first to step and have shorter lifespans than proteins. In this way, you can see gene expression levels increase and decrease with far

more precision. Altogether, using mass cytometry and PLAYR technology allows for the quantification of over 40 proteins and mRNAs at the same time in single cells. The nuanced changes and differences of gene expression in specific cell types can be analyzed. Hopefully this will make the identification of cell-type specific responses and drug targeting a far more straightforward process.

1.5 Use PLAYR system with mass cytometry to detect gene expression in PBMCs infected with flavivirus (Zika virus and Dengue virus)

Flavivirus is a kind of viruses in the family Flaviviridae. This genus includes the West Nile virus, Dengue virus, tick-borne encephalitis virus, yellow fever virus, Zika virus and several other viruses which may cause encephalitis, ([31] Shi, P-Y (editor) 2012: Caister Academic Press. ISBN 978-1-904455-92-9.) and also insect-specific flaviviruses (ISFs) such as cell fusing agent virus (CFAV), Palm Creek virus (PCV), and Parramatta River virus (PaRV). ([32] McLean, B.J. et al. 2015: Virology. 486: 272.) Take West Nile virus as an example, most people (8 out of 10) infected with West Nile virus do not develop any symptoms. About 1 in 5 people who are infected develop a fever with other symptoms such as headache, body aches, joint pains, vomiting, diarrhea, or rash. Most people with West Nile virus disease recover completely, but fatigue and weakness can last for weeks or months. About 1 in 150 people who are infected develop a severe illness affecting the central nervous system such as encephalitis (inflammation of the brain) or meningitis (inflammation of the membranes that surround the brain and spinal cord). The virus can cross the blood-brain barrier by increasing endothelial cell permeability, by the breakdown of endothelial cells or through the transport of infected immune cells. As single-stranded RNA viruses, flavivirus can infect a wide range of target cells, such as monocytes, dendritic cells and macrophages. The virion entry is initiated after the envelope protein, protein E, engages

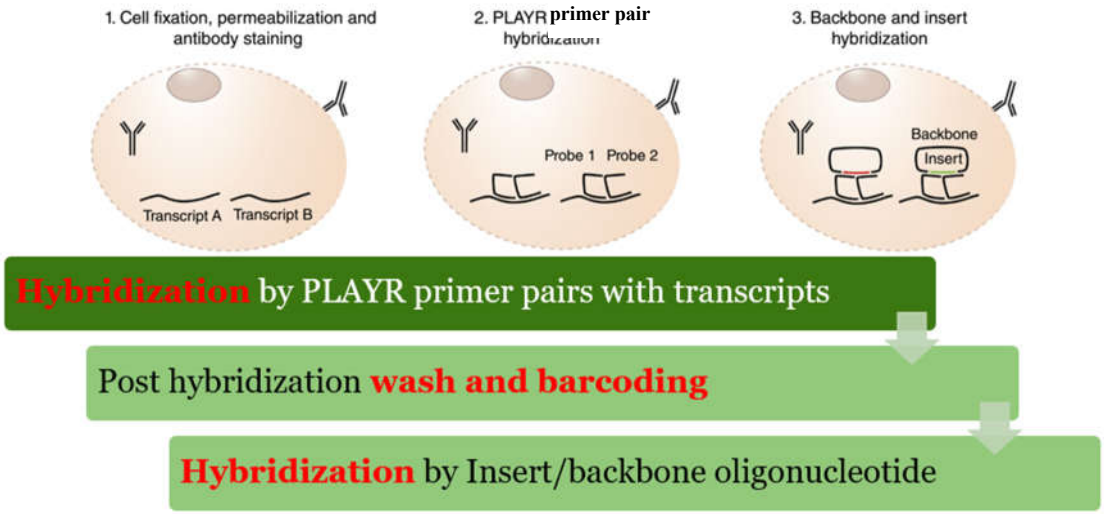
cellular receptors, followed by receptor-mediated endocytosis of the virus. The viral RNA is translated into a single polyprotein at the endoplasmic reticulum and cleaved into mature proteins. Some non-structural proteins of the virus will help making the mature viral proteins. Also, lots of full-length positive ssRNAs will be produced in infected cells. So the envelope proteins, the non-structural proteins and the ssRNAs can be served as our detection targets in infected cells. As for gene expression, innate immune responses to flavivirus includes the activations of IRF3-dependent genes, pro-inflammatory cytokine genes and Type I IFN genes. ([33] Suthar M.S. et al. 2013: Nat. Rev. Microbiology. 11: 115) Given the great differences in susceptibility of different patient populations to flavivirus, we may be able to explain the variations of symptoms in human populations by identifying the differences in the activation of the responsible gene expressions in human PBMCs infected with flavivirus on both transcription and translation levels. This can be done with our PLAYR system together with mass cytometry. For this purpose, we applied this technology on the PBMCs infected with flavivirus (Zika virus and Dengue virus) and found the increase of the transcription and translation of *IL8, IFNB1, IL1, TNF, and CXCL10* in monocytes, indicating the activation of pro-inflammatory cytokine gene pathway and Type I IFN gene pathway.

Chapter Two Results

2.1 Summarize the new PLAYR system with mass cytometry

PLAYR uses the concept of proximity ligation ([34] Fredriksson, S. et al. 2002: Nat. Biotechnol. 20, 473. [35] Söderberg, O. et al. 2006: Nat. Methods 3, 995.) to detect individual transcripts in single cells (Figure 1) and is compatible with immunostaining. We designed pairs of DNA oligonucleotide probes to hybridize to two adjacent regions of target transcripts in previously fixed and permeabilized cells. Each probe should

include one region to selectively hybridize to its cognate target RNA sequence and another region to act as a template for the binding and circularization of two additional oligonucleotides (termed backbone and insert). Backbone and insert oligonucleotides can be ligated once hybridized to an adjacent probe pair, and amplified through rolling circle amplification by phi29 polymerase to produce concatenated complementary copies of the original circle. ([36] Lizardi, P.M. et al. 1998: Nat. Genet. 19, 225.) The amplified product of any probe pair can then be detected using an oligonucleotide labeled with a kind of metal. The cells will then go through the CyTOF machine one by one to let the amount of each kind of metals be measured, which can demonstrate the amount of each kind of transcripts that is labeled by the corresponding type of metals. The detailed steps are shown in Figure 4.



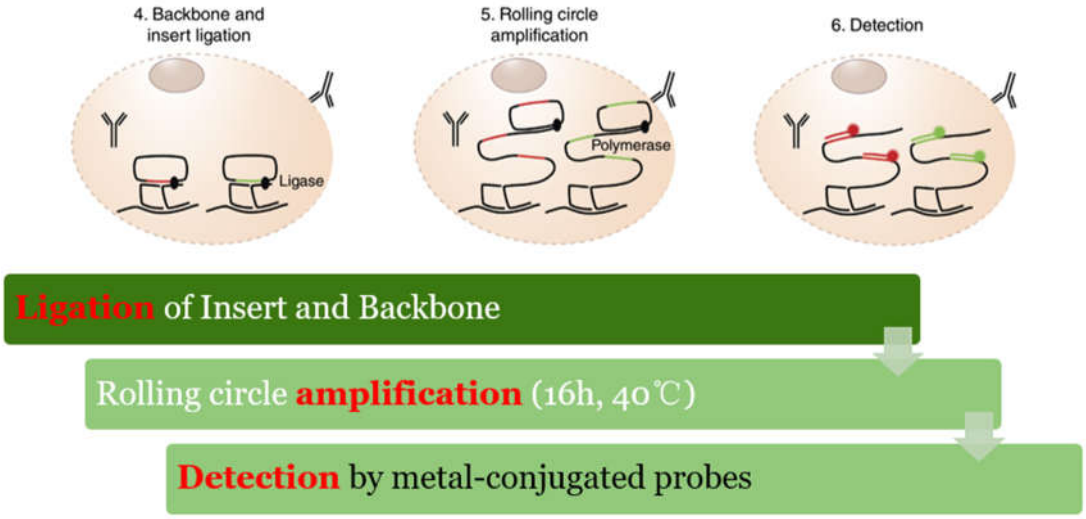
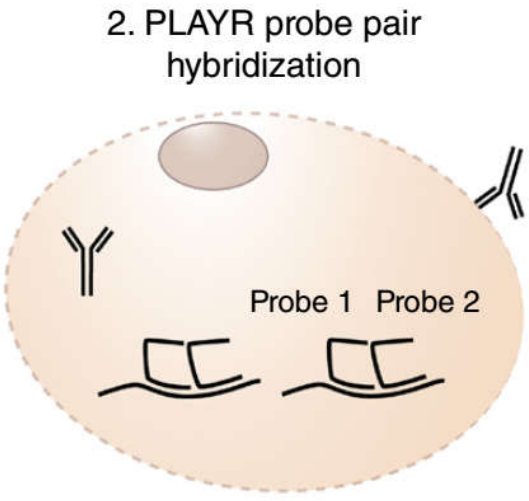


Figure 4 Detailed steps of our PLAYR system compatible with mass cytometry

Because both PLAYR primers must be hybridized independently so that the ligation and rolling circle amplification can take place, the approach will produce low background binding and high specificity. Nonspecific, off-target binding of single primers does not result in a signal. This is different from FISH approaches, in which nonspecific binding of individual primers is possible lead to background signals that cannot be distinguished from specific RNA signals by flow cytometry. PLAYR can be multiplexed through the use of designed oligonucleotides with different insert regions that act as cognate barcodes for given transcripts. Different insert sequences can be designed to have the same melting temperatures and base compositions and form rolling circle amplification products with similar efficiency. ([25] Frei, A.P. et al. 2016: Nat Methods.13(3), 269.) To ensure that rolling circle amplification products uniquely barcode a particular transcript, the insert sequences do not have common substrings longer than 4 bases.

2.2 Design the primer pairs

We used the open-source, user-friendly R software package for the rapid design of PLAYR primer pairs invented by researchers from Stanford University (Figure 5a) that includes a graphical user interface. ([25] Frei, A.P. et al. 2016: Nat Methods.13(3), 269.) Candidate primer pairs with similar thermodynamic properties are produced using Primer3 (Figure 5d), and the application displays their location along the target transcript (Figure 5e). BLAST matches to other transcripts or repetitive genomic sequences are also given (Figure 5 b, c). For each gene, the user can then use these features to guide manual selection of the best primer pairs, in combination with one of the PLAYR insert systems for multiplexing. The primer pairs that have higher summed Primer3 scores are favored. The software then outputs complete probe sequences for detecting transcripts of interest (a list of all probe and backbone-insert system sequences used in this study is presented in Supplementary Table 1).



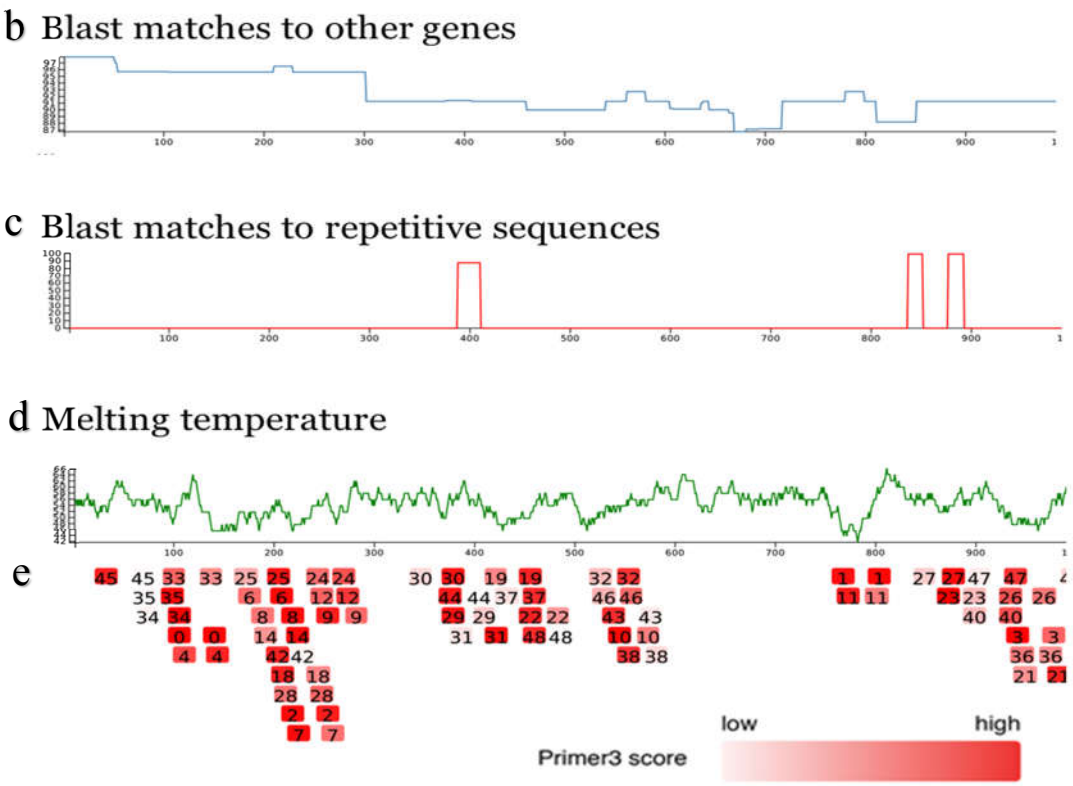


Figure 5 Graphical display of the PLAYRDesign software tool for user-friendly design of PLAYR probe pairs

Each potential primer is represented by a red rectangle. The Primer3 score of each primer is represented by a color gradient from light pink to red, where red primers have higher scores and are preferred over light red primers. The position of primers along the transcript is represented together with sequence features that can guide primer selection. Different graphs represent: maximum sequence identity of BLAST matches to other transcripts (blue top), maximum sequence identity of BLAST matches to a database of repetitive sequences (red); predicted melting temperature in a window of 20 residues (green). The actual melting temperature of primers is independently calculated by Primer3, while the purpose of the green graph is to give an indication on whether certain regions of the transcript have a melting temperature that is too low or too high to be amenable for probe design. Blue and red graphs represent sequence features that are not considered in the scoring of Primer3 primers.

2.3 Conjugate the detection probes with metals and verify the product

To be able to use CyTOF to detect the amount of transcripts in the cells, we need to synthesis the detection probes that are ligated with metals (Figure 6a). The protocol is similar with the steps that synthesis the metal-conjugated antibodies. ([37] Bandura, D.R. et al. 2009: Anal. Chem. 81, 6813. [38] Lou, X. et al. 2007: Chem. Int. Ed. Engl. 46, 6111.) The original detection probes synthesized should have thiol modification on their five prime. In the process of the synthesis, the X8 polymers with several S-S thiol groups will first be loaded with one kind of metals. At the same time, S-S thiol modification on the detection probes will need to be reduced by TCEP and the reduced probes will be collected through ethanol precipitation. Then the reduced and collected detection probes will be ligated with the metal-loaded polymers through incubation, followed by the addition of TCEP to reduce the unconjugated detection probes. Finally, the conjugates will be purified through 30-kDa centrifugal filter units. Typically, one polymer can be loaded with 10-20 metals. One reduced detection probe (~6.3kD) can be conjugated with one polymer (10~12kD), making the conjugates weigh about 16~18kD, in contrast with the unconjugated detection probes (~6.6kD). The success of the conjugation can be verified from the weight difference shown on the polyacrylamide gel electrophoresis (PAGE) (Figure 6b). We deduced that the phenomenon that the signal on the PAGE will look blurred is because of the varying number of metals loaded on the polymers and the different weight of the polymers. We cannot run the sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) since the ligated conjugates might be reduced by sodium dodecyl sulfate.

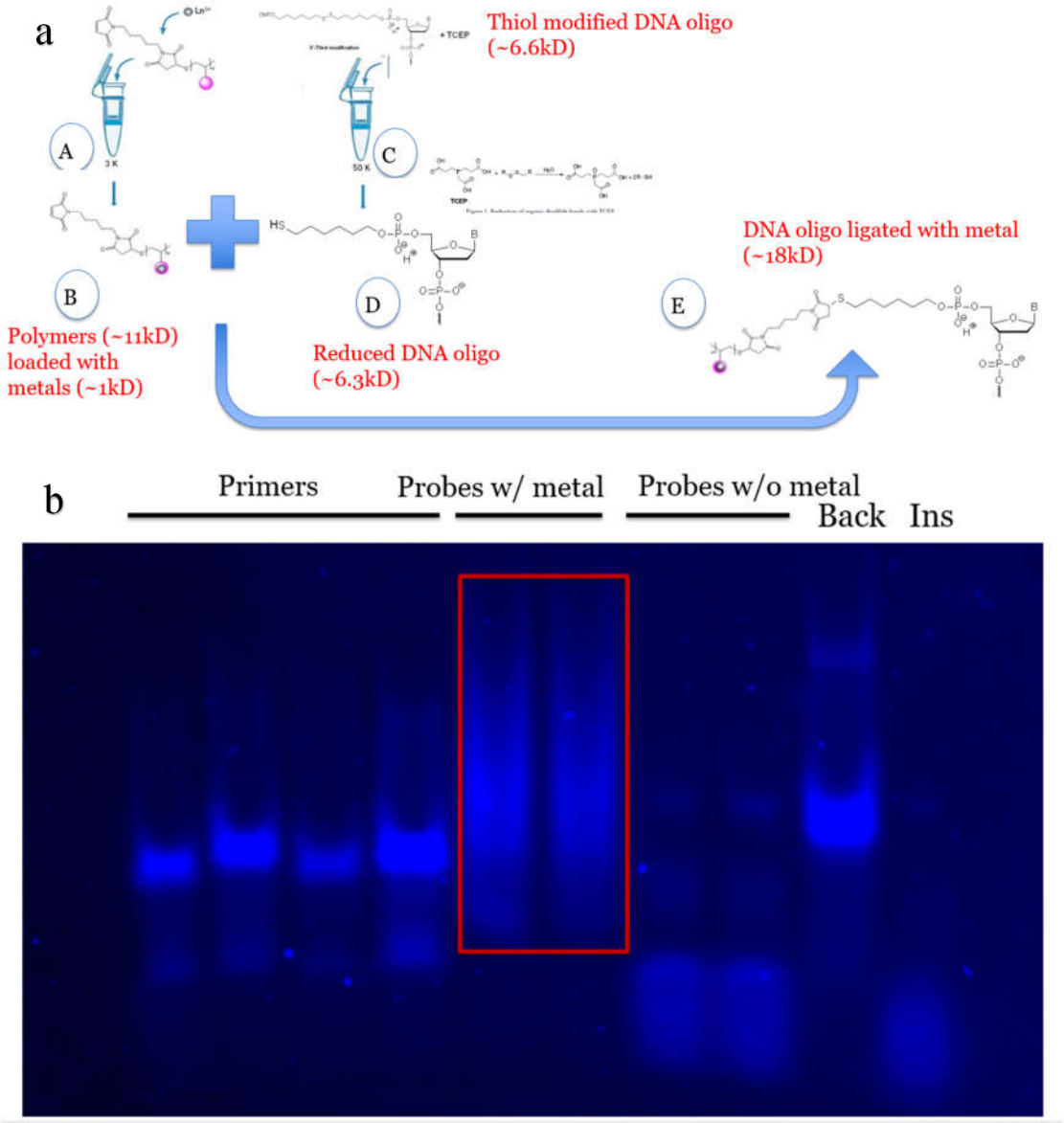


Figure 6 Conjugate the detection probes with metals and verify the product

a) The new procedures for ligating the detection probes with the metals. The metals are loaded on the polymers (~11kD). At the same time, the 5' thiol modified detection probes (~6.6kD), which have been synthesized by professional facilities, need to be reduced by TCEP, leaving a bare thiol group at the end the thiol modification part. Then the polymers (~12kD) and the reduced detection probes (~6.3kD) can be ligated with each other through oxidation to produce the metal-conjugated detection probes (16~18kD).

b) The PAGE verification by comparing the weight of the product with other oligos.

2.4 Validate the effect of the PLAYR protocol on surface and intracellular antibody staining

Before we carried out the PLAYR system on the cells to detect RNAs, we had to validate the antibody staining part of the PLAYR protocol, especially the surface antibody staining part. According to the original PLAYR protocol from Stanford researchers ([25] Frei, A.P. et al. 2016: Nat Methods.13(3), 269.), which we call "PLAYR staining", the cells were fixed first and the surface staining is combined with intracellular staining. It is reasonable to do the fixation at the very beginning and then immediately add the reagents to protect RNAs in the cells together with the permeable reagents and antibodies. However, conventionally, we do surface staining on live cells without fixation, hence "surface staining of live cells" is often used. ([30] Bendall, S.C. et al. 2011: Science 332, 687.) As a result, we needed to make sure that the surface staining results from the experiments using "PLAYR staining" was comparable with the conventional "surface staining of live cells", which is why we added a group that used the reagents according to the original PLAYR protocol, but separated the intracellular staining from the surface staining. We call it "PLAYR staining of live cells" (Figure 7a).

From the figures showing our gating strategy, it can be seen that almost all the innate immune cell populations can be clearly identified with the "PLAYR staining" procedures. The specific cell populations, such as the CD3- CD19- CD14- cells are identified from total single live PBMCs using the surface antibody staining (Figure 7b). By comparing the frequencies of this cell population between the group "surface staining of live cell" and the group "PLAYR staining of live cells", we did not find

significant variations, which could indicate that the surface staining was not affected significantly by combining the two staining steps (Figure 7c). In addition, from the comparison of our conventional "surface staining of live cells" and "PLAYR staining of live cells", we did not find significant effect of the reagents from Stanford researchers on surface staining, validating that the unique reagents from Stanford, mainly for the protection of RNA in the cells, can be used to achieve similar surface staining effect.

Furthermore, we also validated the efficacy of intracellular antibody staining with "PLAYR staining" protocol. The baseline expression of IL6 was investigated in CD14+ monocytes (Figure 7d). Using viSNE plot, in the cell populations stained with PLAYR staining, the cell population that is highly expressing CD14, is also highly expressing IL6, which is consistent with the notion that the baseline expression level of IL6 is relatively high in monocytes.

We can reach the conclusion that the two staining steps can be combined without significantly affecting intracellular staining.

a

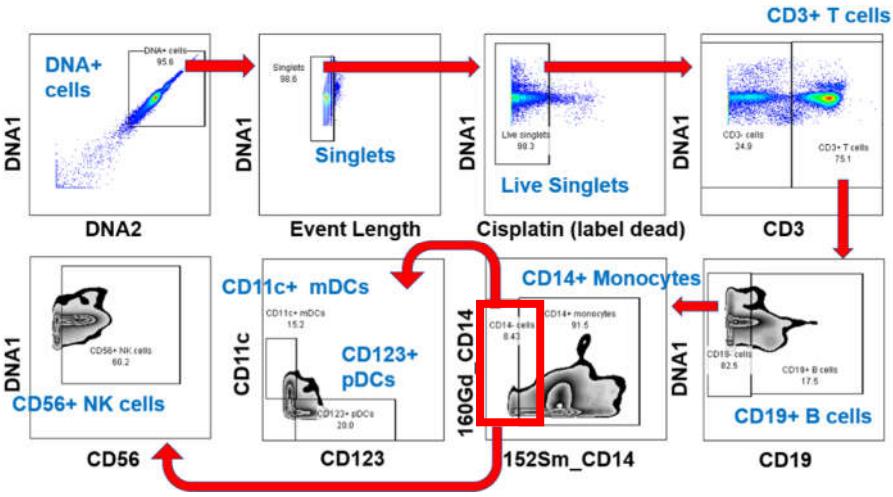
Surface staining of live cells (conventional): Surf Ab → fixation → perm → intra Ab

PLAYR staining (Stanford protocol): Fixation → (perm)1 + surf Ab + intra Ab → (perm)2

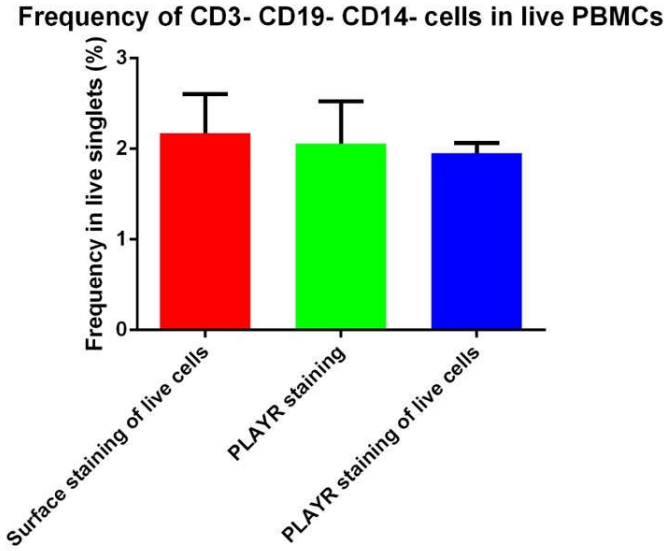
PLAYR staining of live cells (separated staining steps with unique reagents):

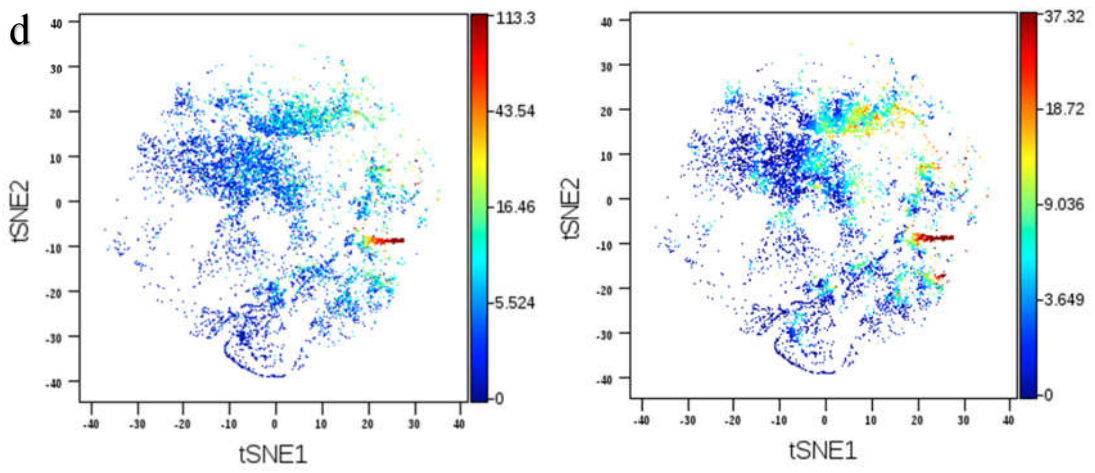
Surf Ab → Fixation → (perm)1 + intra Ab → (perm)2

b


DNA+ cells
DNA1
DNA2
Singlets
Event Length
Live Singlets
Cisplatin (label dead)
CD3+ T cells
CD3
CD19+ B cells
CD19
CD14+ Monocytes
152Sm_CD14
160Gd_CD14
CD11c+ mDCs
CD11c
CD123+ pDCs
CD123
CD56+ NK cells
CD56

c


Frequency of CD3- CD19- CD14- cells in live PBMCs
Frequency in live singlets (%)
0
1
2
3
Surface staining of live cells
PLAYR staining
PLAYR staining of live cells



Expression of CD14 Expression of IL6

Figure 7 Validate the effect of the PLAYR protocol on surface and intracellular antibody staining

a) The groups design for verifying the antibody staining part of the PLAYR protocol. "The surface staining of live cells" group is using the conventional mass cytometry staining procedures, which separates the surface staining from the intracellular staining. "The PLAYR staining" group is the antibody staining protocol from Stanford, which combines intracellular staining and surface staining after cell fixation and uses the unique reagents. "PLAYR staining of live cells" group is using the reagents from Stanford, but separating the surface staining from intracellular staining.

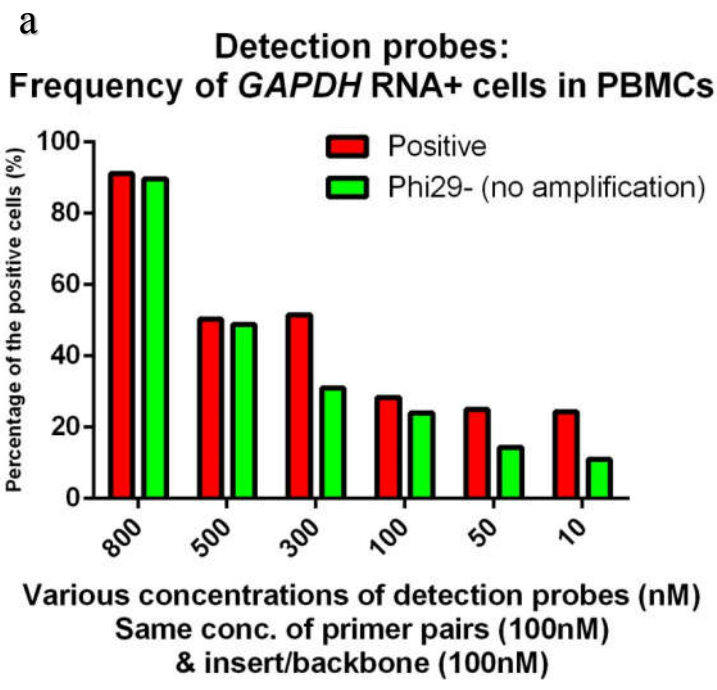
b) Gating strategy of the PBMCs after conducting the different procedures. Different immune cell population are identified using surface antibody staining.

c) The quantification of the specific cell population (CD3- CD19- CD14- cells)with specific surface antibody staining.

d) Intracellular cytokine protein detection in all the human PBMCs. Detection of IL6 in human PBMCs (left). Detection of CD14 on human PBMCs (right).

2.5 Detect *GAPDH* mRNA by CyTOF (determine the concentration of primer pairs, Insert/Backbone oligos and detection probes)

According to the original PLAYR protocol from Stanford researchers, all the concentration of the primer pairs and Insert/Backbone oligos were 100nM in one million cells per sample. This concentration is suitable for the PLAYR protocol that is compatible with flow cytometry. However, the detection probes used in PLAYR system with flow cytometry should be different from the one with mass cytometry. As a result, we first experimented a series of concentrations of detection probes to detect *GAPDH* mRNA in PBMCs in PLAYR system with CyTOF. We found that the highest signal-background ratio can be achieved with 300nM of detection probes in one million cells per sample (Figure 8a). Then we verified a few various concentrations of primer pairs and Insert/Backbone based on the original protocol, leading to the establishment of using 500nM of primer pairs and 100nM of Insert/Backbone in the new PLARY protocol (Figure 8b).



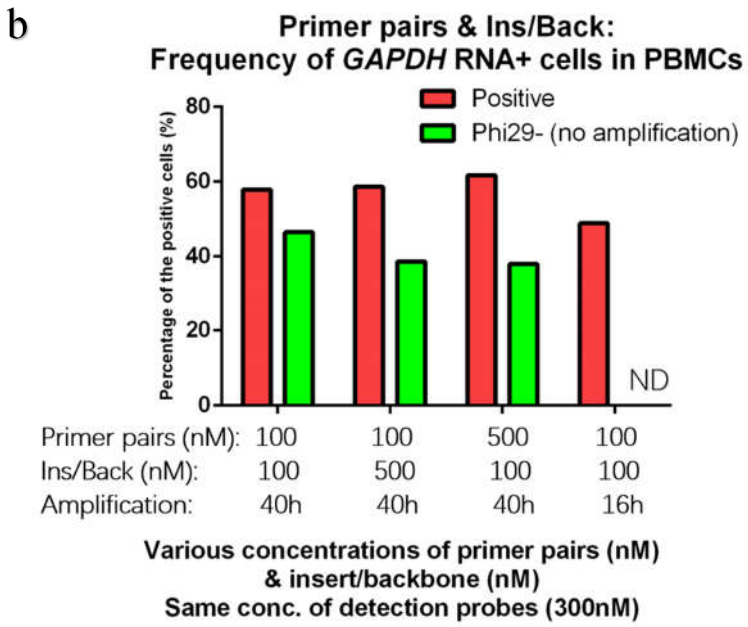


Figure 8 Detect *GAPDH* mRNA by CyTOF (determine the concentration of primer pairs, Insert/Backbone oligos and detection probes)

a) Detect *GAPDH* RNA in human PBMCs using various concentrations of the detection probes.

b) Detect *GAPDH* RNA in human PBMCs using various groups of concentrations of the primer pairs and the Insert/ Backbone oligos. (ND: No Data)

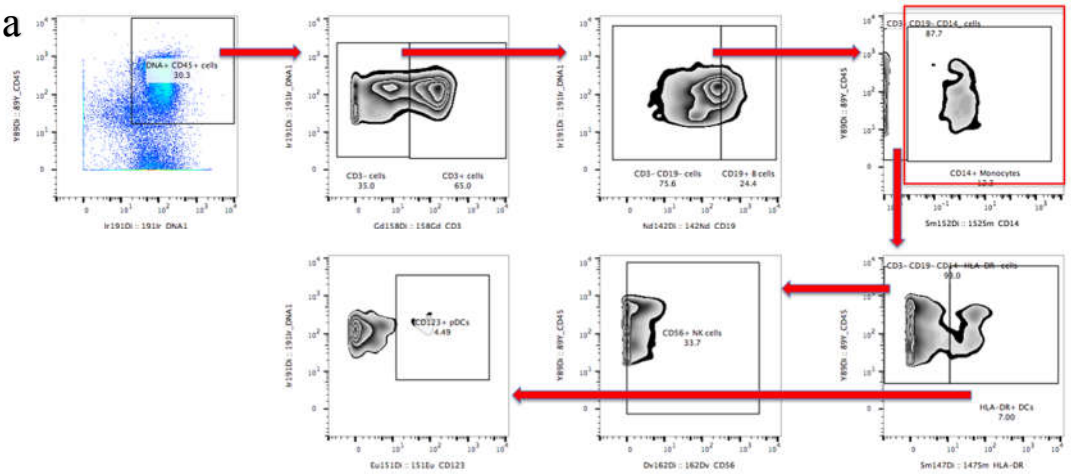
2.6 Detect several proteins and RNA simultaneously by CyTOF

2.6.1 Detect mRNA and proteins of the cytokines in monocytes from PBMCs stimulated with R848

To apply our technology, we stimulated human PBMCs with R848 (Resiquimod). R848 is a ligand for TLR7/8. Activation of TLR7/8 can trigger the recruitment of the adapter molecule MYD88 leading to the activation of NF-κB and other transcription factors

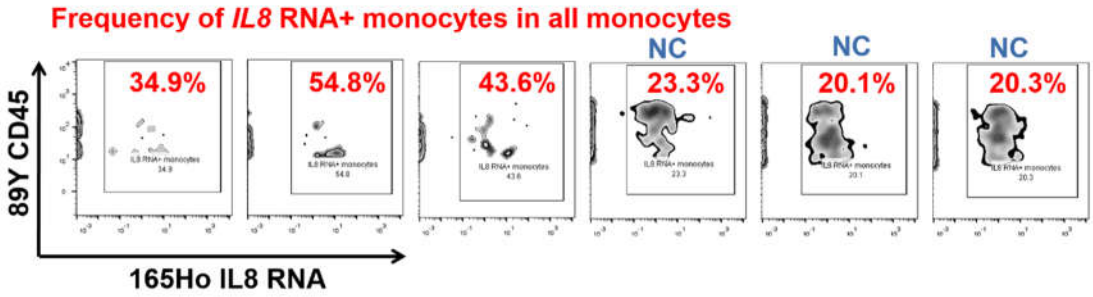
and the production of proinflammatory cytokines and chemokines, like TNFa, IL-8, CXCL10 and IL-1. Also, MYD88 recruited by TLR7/8 can result in the activation of IRF7 and Type I IFN genes, such as *IFNa*. TLR7/8 is highly expressed on antigen presenting cells, ([39] Gorden, K.B. et al. 2005: J Immunol. 174(3), 1259.) which include dendritic cells (DCs), macrophages and monocytes, leading to the expression of the proinflammatory cytokines and chemokines and Type I IFNs in those cells. With our PLAYR system that is compatible with CyTOF, we are able to use epitope detection to identify multiple cell subsets in live PBMCs (Figure 9a). Then we are able to detect the transcripts of certain cytokine genes (*TNFa, IL-8, IL-1, and IFNB1*). As we expected, with R848 stimulation, the expressions of all the cytokines were upregulated on transcription level (Figure 9b, 9c). In accordance, the expression of certain cytokine genes, *IL1B, IL8, TNFa*, were increased on protein level (Figure 9c). On the contrary, the expression of *IFNB1*, did not show obvious upregulation on translation level (Figure 9c). In addition, from the viSNE plots, in the PBMCs stimulated with R848, the cells that are highly expressing *IL8, IFNB1, IL1B* genes on RNA level and protein level can be correspondent with each other for each gene expressions (Figure 9d), which is the same as the previous studies. ([39] Gorden, K.B. et al. 2005: J Immunol. 174(3), 1259.) The results indicates that our technology can indeed enable CyTOF to achieve highly multiplexed protein and transcript detection in single PBMCs. Moreover, the increase in RNA expression of most genes can be correspondent with the protein expressions, which to some extent validated our method. Finally, our PLAYR technology with CyTOF is able to detect the increase of the expressions of some genes that cannot be shown by traditional mass cytometry detection, making PLAYR a useful complement to the cutting-edge CyTOF system.

a

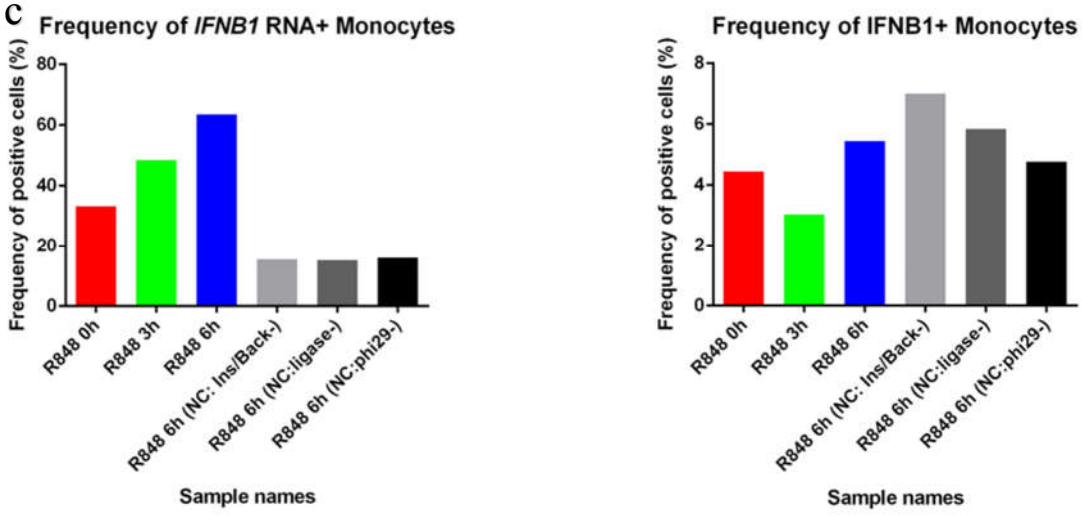


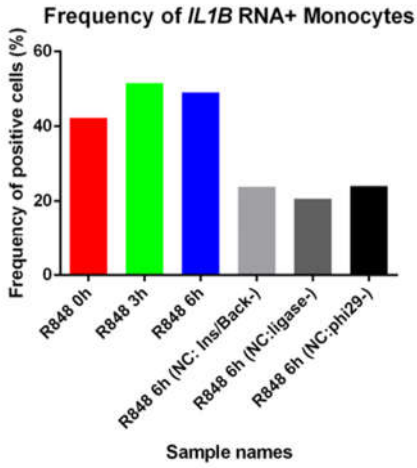
b

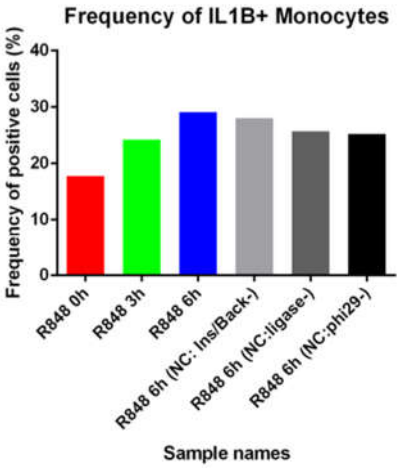
Sample	R848 0h/Amplify+	R848 3h/Amplify+	R848 6h/Amplify+	NC: R848 6h/InsBack-	NC: R848 6h/Ligase-	NC: R848 6h/Phi29-
R848	0h (BFA 3h)	3h (BFA 3h)	6h (BFA 3h)	6h (BFA 3h)	6h (BFA 3h)	6h (BFA 3h)
Insert/Backbone	+	+	+	-	+	+
Ligase	+	+	+	+	-	+
Phi29	+	+	+	+	+	-

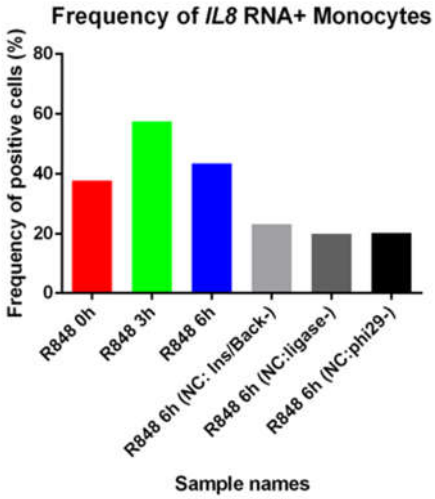


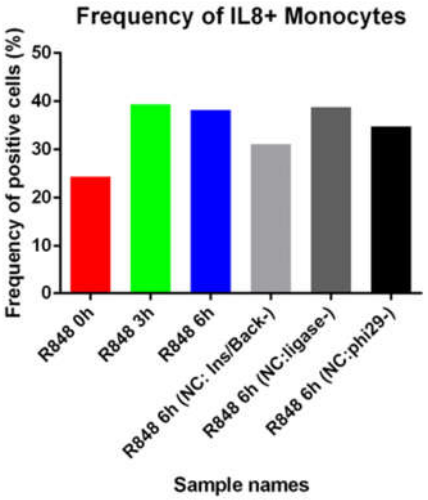
c

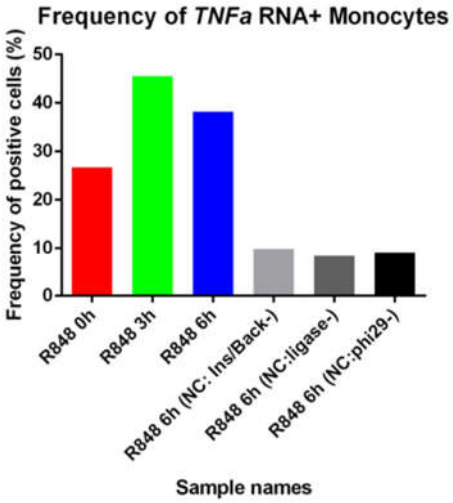


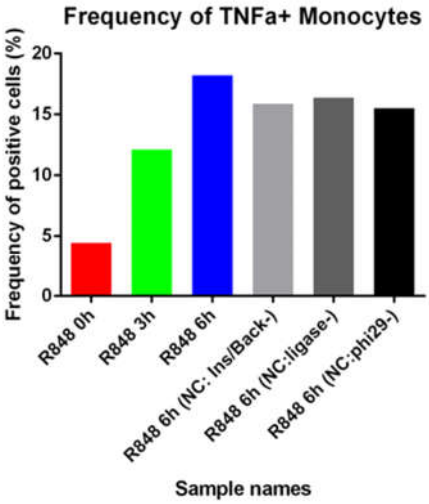

Frequency of IL1B RNA+ Monocytes
Frequency of positive cells (%)
60
40
20
0
R848 0h
R848 3h
R848 6h
R848 6h (NC: Ins/Back-)
R848 6h (NC:ligase-)
R848 6h (NC:phi29-)
Sample names


Frequency of IL1B+ Monocytes
Frequency of positive cells (%)
40
30
20
10
0
R848 0h
R848 3h
R848 6h
R848 6h (NC: Ins/Back-)
R848 6h (NC:ligase-)
R848 6h (NC:phi29-)
Sample names


Frequency of IL8 RNA+ Monocytes
Frequency of positive cells (%)
80
60
40
20
0
R848 0h
R848 3h
R848 6h
R848 6h (NC: Ins/Back-)
R848 6h (NC:ligase-)
R848 6h (NC:phi29-)
Sample names


Frequency of IL8+ Monocytes
Frequency of positive cells (%)
50
40
30
20
10
0
R848 0h
R848 3h
R848 6h
R848 6h (NC: Ins/Back-)
R848 6h (NC:ligase-)
R848 6h (NC:phi29-)
Sample names


Frequency of TNFa RNA+ Monocytes
Frequency of positive cells (%)
50
40
30
20
10
0
R848 0h
R848 3h
R848 6h
R848 6h (NC: Ins/Back-)
R848 6h (NC:ligase-)
R848 6h (NC:phi29-)
Sample names


Frequency of TNFa+ Monocytes
Frequency of positive cells (%)
20
15
10
5
0
R848 0h
R848 3h
R848 6h
R848 6h (NC: Ins/Back-)
R848 6h (NC:ligase-)
R848 6h (NC:phi29-)
Sample names

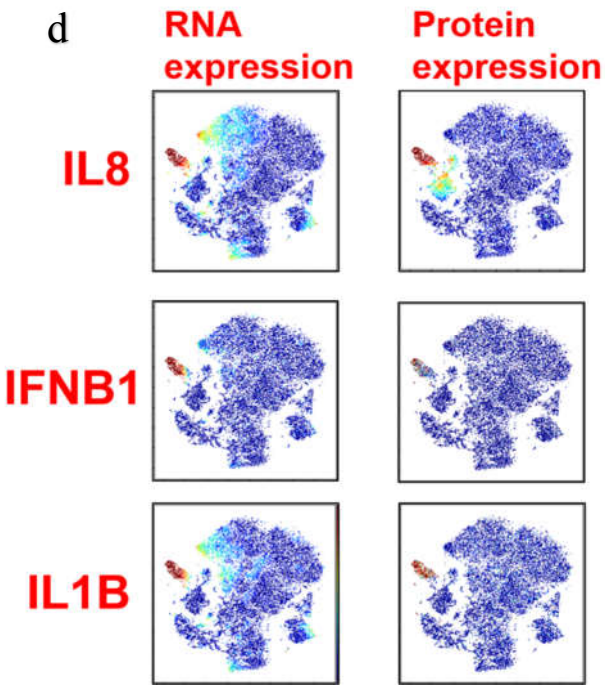


Figure 9 Detect mRNA and proteins of the cytokines in monocytes from PBMCs stimulated with R848

a) The gating strategy in the experiments.

b) The experiment design and the frequency of *IL8* RNA+ monocytes in all the monocytes in the PBMCs of each group (NC: Negative Control). The group design (up) and the results (down) are correspondent with each other.

c) Quantification of the frequency of the monocytes positive for *IFNB1, IL1B, IL8, TNFa* RNA and proteins in all the monocytes of each group.

d) The viSNE plot showing the RNA and protein expressions of *IL8, IFNB1, IL1B* genes in the PBMCs stimulated with R848.

2.6.2 Detect mRNA and proteins of the cytokines in monocytes from PBMCs infected with flavivirus (Dengue virus and Zika virus)

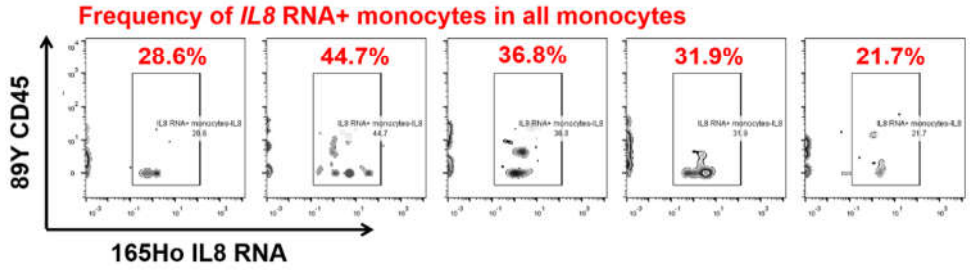
We also applied this technology on the human PBMCs infected with flavivirus (Zika virus and Dengue virus). Flaviviruses are all positive-sense, single-stranded RNA virus, which made them the strong activators of TLR7/8. As a result, similar to the situation of R848 stimulation, activation of TLR7/8 can trigger the recruitment of the adapter molecule MYD88 leading to the activation of NF-κB and other transcription factors and the production of proinflammatory cytokines and chemokines, like TNFa, IL-8, CXCL10 and IL-1. Also, MYD88 recruited by TLR7/8 can result in the activation of IRF7 and Type I IFN genes, such as *IFNa*. ([38] Lou, X. et al. 2007: Chem. Int. Ed. Engl. 46, 6111.) With our PLAYR system that is compatible with CyTOF, we were able to use epitope detection to identify multiple cell subsets in live PBMCs. Then we were able to detect expression of the transcripts of certain cytokines (TNFa, IL-8, CXCL10, IL-1, IFNa). As we expected, after the infection of human PBMCs with Zika virus (MOI = 5 or 10) or Dengue virus (MOI = 5 or 10) for 48 hours, the expressions of all the detected cytokines were upregulated on transcription level in monocytes compared with the Mock group (Figure 10b). For the all the detected cytokines, the transcription level expression of the genes was upregulated with the increase of the concentrations of Dengue virus (Figure 10b). In contrast, in spite of the increase of the concentrations of Zika virus, the transcription level expression of the *IFNB1, IL8, and CXCL10* genes did not show great differences. The *IL1B* and *TNFa* RNA even exhibited decrease in the MOI10 group of Zika virus infected cells compared with MOI5 group (Figure 10b). It is possible that with the increase of the virus concentration, the immune system is exhausted and hence the proinflammatory cytokine gene pathway and Type I IFN gene pathway are gradually inactivated. In consistent with the results of the RNA expression of the cytokines in monocytes, the protein expression of certain cytokine genes, *IL1B, IL8, CXCL10 and TNFa* showed obvious upregulation with the infection of both Dengue virus and Zika virus compared with the Mock group. The protein level of *IL8* expression is upregulated with the increase of the concentration of Dengue virus, which is consistent with the results obtained from RNA detection.

However, the protein level of the expressions of *IL1B, IFNB1, THFa and CXCL10* did not show similar upregulation with the increase of the Dengue virus concentration (Figure 10b). This result may be changed after the improvement of our technology. On the other hand, this result might demonstrate the strength of our technology since we can detect the transcriptional level upregulation of certain cytokine genes with cannot be detected with mere protein detection of previous mass cytometry technology. More importantly, the viSNE plots showed that in the PBMCs infected with Dengue virus, the cells, in which the pro-inflammatory cytokines are activated, had higher expressions of *DENV* RNA inside compared with the other cells (Figure 10c). This result indicated that our new PLAYR technology can achieve the measurement of the amount of virus RNA in single cells and the identification of the cell populations that infected with the virus.

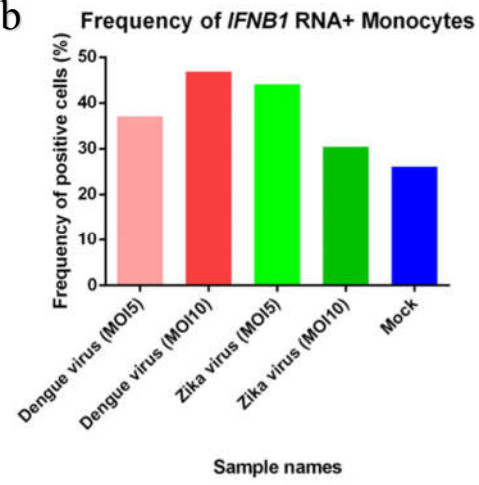
In conclusion, similar to the previous results, virus infection experiments indicated that our technology can indeed enable CyTOF to achieve highly multiplexed protein and transcript detection in single PBMCs. Moreover, the increase in RNA expression of some genes can be correspondent with the protein expressions, which to some extent validated our method. Most importantly, by applying our PLAYR system with CyTOF to real in vitro virus infection experiments, we made a successful start on using our technology to characterize various gene expression patterns in single human PBMCs infected with flavivirus in vitro. This is the first time that the PLAYR technology has been applied to the pathology research field. In the future, we may be able to move forward to apply our technology to identify the gene expression patterns in the PBMCs from virus-infected patients so that we may finally be able to explain the differential susceptibility to flavivirus virus among different people.

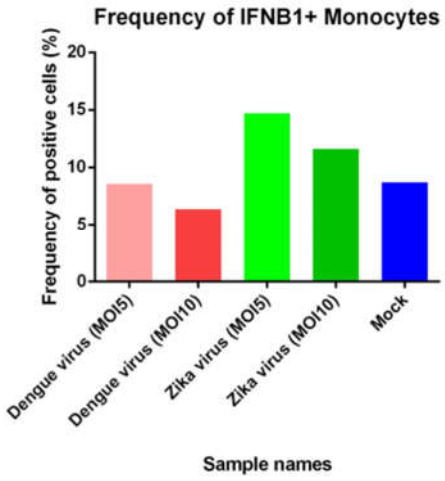
a

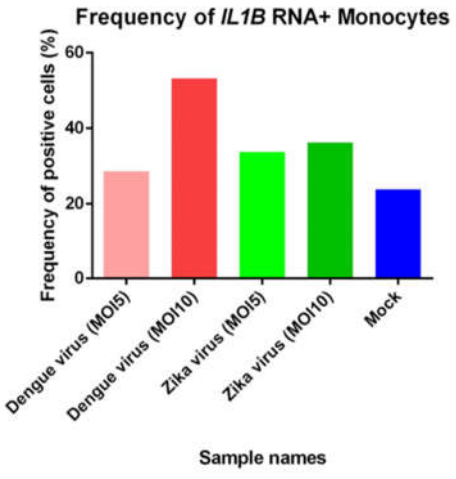
Sample	Dengue virus (MOI=5)	Dengue virus (MOI=10)	Zika virus (MOI=5)	Zika virus (MOI=10)	Mock
Durati-on	48h (BFA 3h)				

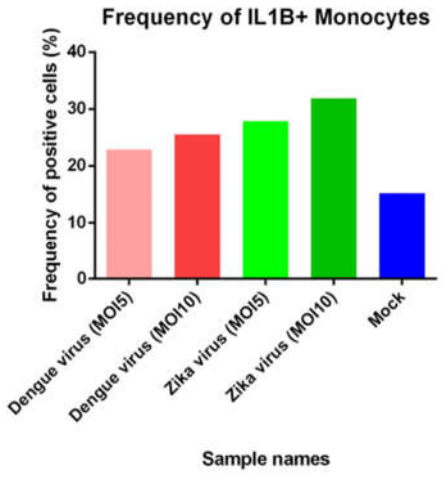


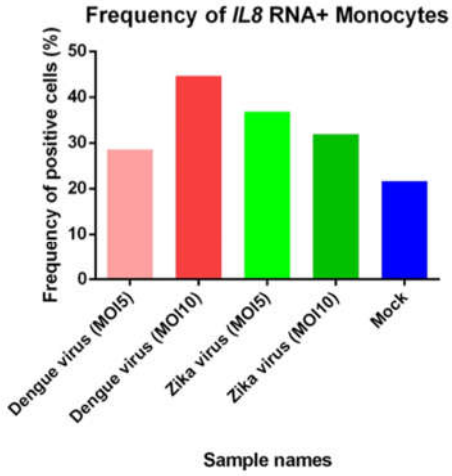
b

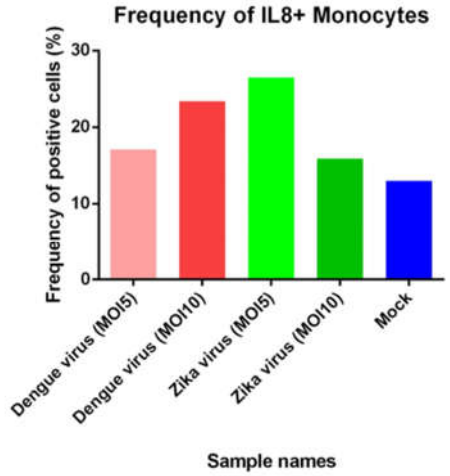


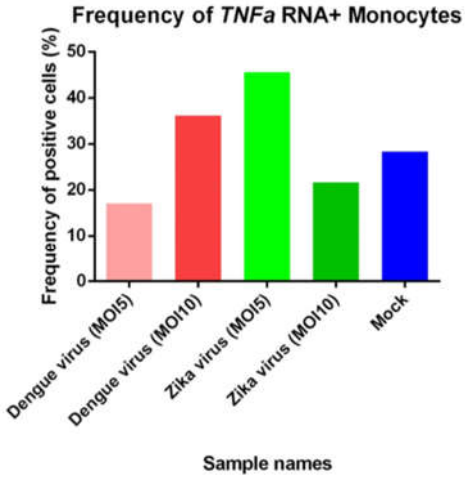


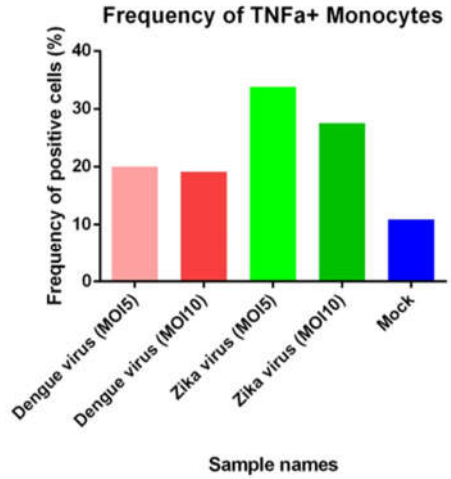


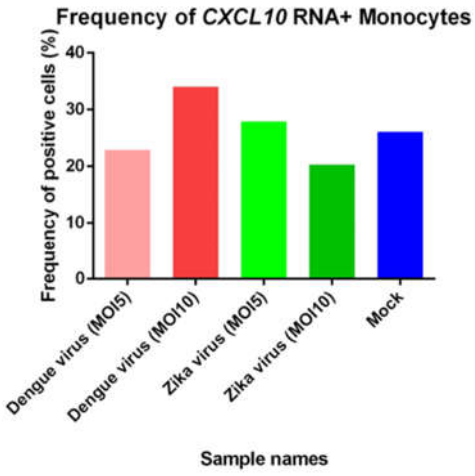


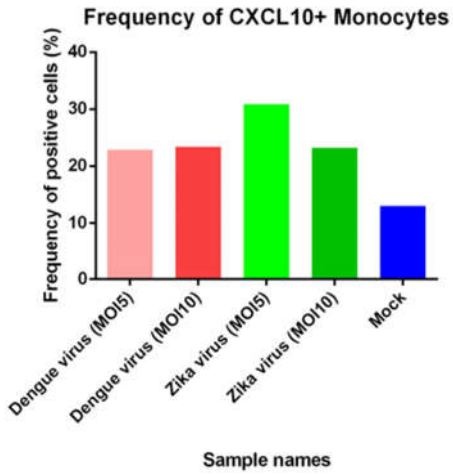

Frequency of *IL8* RNA+ Monocytes
Frequency of positive cells (%)
0
10
20
30
40
50
Dengue virus (MOI5)
Dengue virus (MOI10)
Zika virus (MOI5)
Zika virus (MOI10)
Mock
Sample names


Frequency of IL8+ Monocytes
Frequency of positive cells (%)
0
10
20
30
Dengue virus (MOI5)
Dengue virus (MOI10)
Zika virus (MOI5)
Zika virus (MOI10)
Mock
Sample names


Frequency of *TNFa* RNA+ Monocytes
Frequency of positive cells (%)
0
10
20
30
40
50
Dengue virus (MOI5)
Dengue virus (MOI10)
Zika virus (MOI5)
Zika virus (MOI10)
Mock
Sample names


Frequency of TNFa+ Monocytes
Frequency of positive cells (%)
0
10
20
30
40
Dengue virus (MOI5)
Dengue virus (MOI10)
Zika virus (MOI5)
Zika virus (MOI10)
Mock
Sample names


Frequency of *CXCL10* RNA+ Monocytes
Frequency of positive cells (%)
0
10
20
30
40
Dengue virus (MOI5)
Dengue virus (MOI10)
Zika virus (MOI5)
Zika virus (MOI10)
Mock
Sample names


Frequency of CXCL10+ Monocytes
Frequency of positive cells (%)
0
10
20
30
40
Dengue virus (MOI5)
Dengue virus (MOI10)
Zika virus (MOI5)
Zika virus (MOI10)
Mock
Sample names

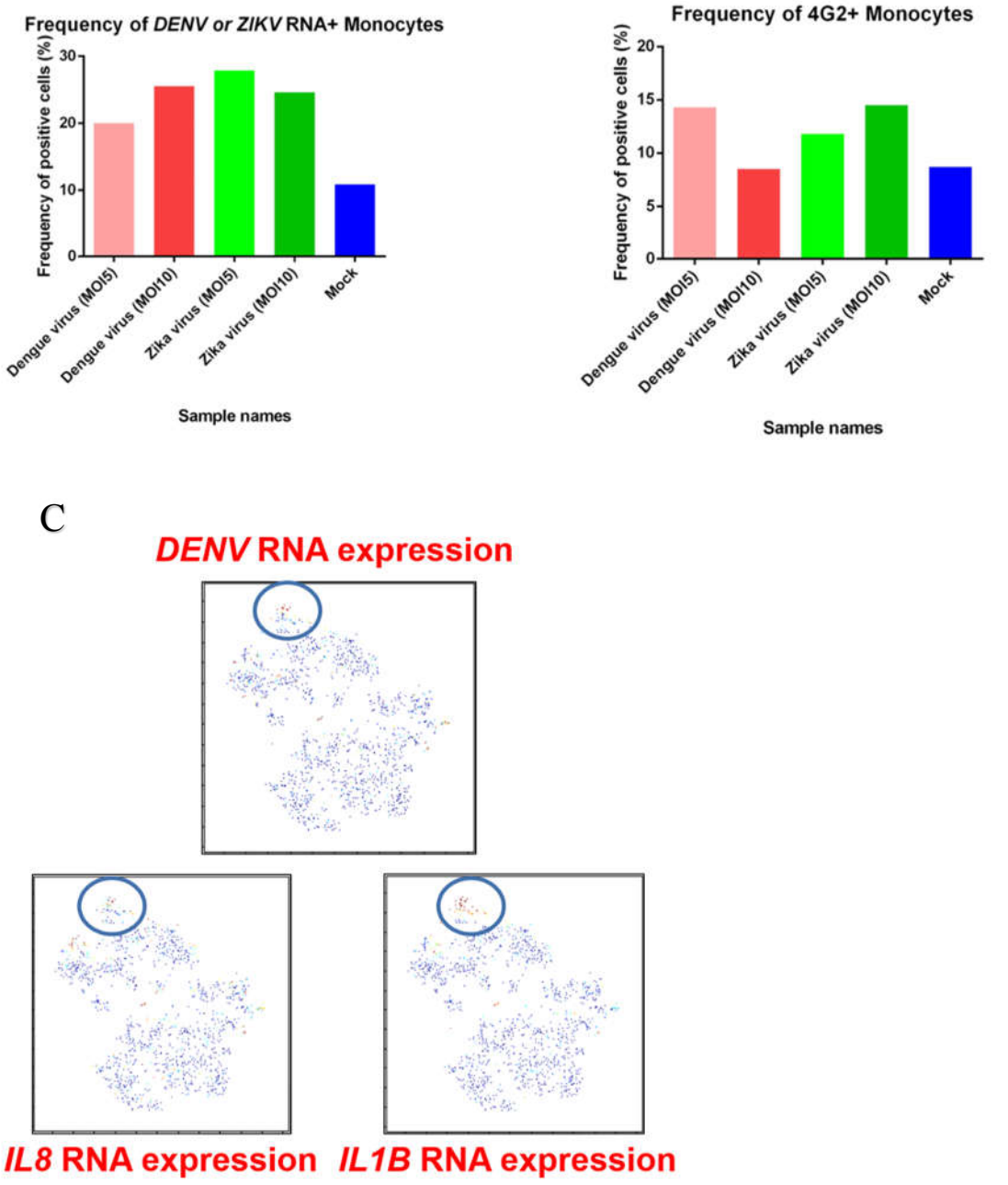


Figure 10 Detect mRNA and proteins of the cytokines in monocytes from PBMCs infected with flavivirus (Zika virus MOI=10 and 5, Dengue virus MOI=10 and 5)

a) The experiment design and the frequency of *IL8* RNA+ monocytes in all the monocytes in the PBMCs of each group. The group design (up) and the results (down) are correspondent with each other.

b) Quantification of the frequency of the monocytes positive for *IFNB1, IL1B, IL8, TNFa, CXCL10 and DENV/ZIKV* RNA and proteins(4G2 is part of the E protein of the virus) in all the monocytes of each group.

c) The viSNE plot showing the *DENV* RNA and *IL8, IL1B* RNA in the PBMCs infected with Dengue virus (MOI=10).

Chapter Three Discussion

In this study, we first developed a new PLAYR system that is compatible with mass cytometry platform. This technology enables CyTOF to achieve highly multiplexed quantification of both mRNAs and proteins in single cells. With this technology, it is possible for us to measure the expression of genes interested on both translation and transcription level. Moreover, we are the first group that applied the PLAYR system on the pathology research, measuring RNA and proteins simultaneously in single cells that are stimulated with ligands or infected with virus. Although the technology may need more improvement and adjustment for different pathology experiments, with the hope of eventually applying this powerful technology to the human subjects, this project has laid a solid foundation for future studies. Due to the huge variation in the physical status of human and the heterogeneity in the healthy human immune system, ([40] Maecher H.T. et al. 2012: Nat. Rev. Immunol. 12: 191. [41] Davis, M. M. 2008: Immunity 29, 835.) it is very important to establish a technology that can collect as many data as possible in one standardized “Human Immunology Project”. As a result, our technology can be a very useful tool in the future “Human Immunology Project”. With more comprehensive understanding of the changing patterns of our immune system in different human populations, during the health related situations like virus infection or cancer or immune compromise, we will be able to more accurately relate certain

patterns of the immune system to certain patients, making contributions to precision medicine.

However, in just one-year research, the PLAYR technology we accomplished was not perfect. Even with the optimized concentrations of the primer pairs, insert and backbone oligos and detection probes, together with the elongated rolling circle amplification period, we still could not achieve high signal-noise ratio. GAPDH gene should be expressed in all PBMCs, but we can only identify about 60% of PBMCs containing GAPDH RNA. Moreover, our negative controls (without Phi29 enzyme, leading to no amplification) should have low frequencies of PBMCs identified as GAPDH RNA positive cells. However, our results showed 40% of live PBMCs as GAPDH RNA positive cells. The possible explanation of the results should be the nonspecific binding of the detection probes. It is also possible that the detection probes were just stuck in the cells. Our next step should be further reducing the concentration of detection probes until we get the proper negative controls. Based on the established concentration of detection probes, we can further adjust the concentrations of the primer pairs and insert and backbone oligos to increase the efficiency of the RNA detection.

Currently CyTOF can not only be used on the cell suspensions, but also be applied on tissues. ([42] Giesen C. et al. 2014: Nat. Methods, 11: 417-422.) To gain spatial information, researchers at University of Zürich have combined immunohistochemical and immunocytochemical methods with high-resolution laser ablation to CyTOF. This approach made possible the simultaneous imaging of 32 proteins and protein modifications at subcellular resolution on tissue samples. Imaging mass cytometry was applied to human breast cancer samples, allowing delineation of cell subpopulations and cell-cell interactions and emphasizing tumor heterogeneity. Imaging mass cytometry can complement existing imaging approaches. It will make possible basic studies of tissue heterogeneity and function and prompt the transition of medicine

toward personalized molecularly targeted diagnosis and therapies. Our PLAYR technology is likely to be combined with imaging mass cytometry and enable us to detect RNAs and proteins simultaneously on tissue samples. If possible, we will be able to extend imaging mass cytometry to achieve multiplexed detection of mRNA and proteins in tissues. This can allow an increase in the information retrieved from patient samples for biomedical purposes, do detailed studies of tumor biology, and be established as a tool to bridge comprehensive genomic and proteomic tissue research.

Chapter Four Online Content

More information about the PLAYR system, especially the design of the primer pairs, are available by github website https://github.com/nolanlab/PLAYRDesign. This github site provides a centralized location for sharing source code, design files, software and other relevant information so that users can share ideas related to this important RNA measurement technique. The goal is to help share this technology to the larger immunology and mass cytometry community and build a foundation of users that will continue advancing this technology and contribute back to the project. I have propagated this technology to Yale Rheumatology Department and Yale CyTOF Community. Both teams can be consulted on this technology.

Chapter Five Materials and Methods

5.1 PBMC isolation, stimulation and incubation

Draw blood in one tube. For anti-coagulant, use heparin (0.15ml of 1000U/ml stock per 10ml human blood). In 50ml tube, mix 1:1 v/v with PBS. Use PBS to rinse out tubes to

avoid losing sample first and then bring up to the final volume. Prepare a 50ml tube with the same amount of Ficoll-Hypaque (Amersham Cat# 17-1440-03) as that of blood. Gently layer diluted blood over Ficoll-Hypaque to achieve 1:2 Ficoll-to-blood ratio. Then spin at 1800 rpm for 20min at RT (24˚C). Set the brake as off. Then mononuclear leukocytes sediment to the plasma-Ficoll interface. Use a plastic pipette (or vacuum sucker inside a pipette tip), discard plasma to above ~0.5cm of cell layer. Then collect cells with a pipette into a new 50ml tube. After collection of cells, add ~20ml of RPMI. Spin at 1500 rpm for 10min at RT. Set the brake as on. All the spin afterwards will set the brake as on. Discard supernatant. Resuspend in 5ml RPMI and transfer the cells to a 15ml tube. Count number of cells. Prepare cell solution and Trypan Blue dye (which tells if cell is live/dead) in 1:9 ratio. Add 10μl of mixture into each well on a Countess slide. Dilute if cell density is beyond accuracy range of the machine. Take average of live cell counts from both wells. Add RPMI to half the volume of the tube. Spin the tube at 1500 RPM for 10min at RT. Discard supernatant. Resuspend cells in PBMC culture medium and rest for 30min at 37˚C. Spin down cells at 300RCF for 5 minutes. Discard the supernatant. Resuspend cells in RPMI at a density of 1 million/ml.

5.2 Virus infection or stimulation of the incubated PBMCs

5.2.1 R848 stimulation of the incubated PBMCs

For R848 stimulation experiments, incubate about one million cells with 10uM R848 reagent in one well of 48-well plate. Incubate the cells for 6h at 37˚C, 5% CO2. At 3h before the end of the incubation, add Brefeldin A (BioLegend) to block the secretion of the cytokines. At the end of the incubation, add EDTA at the concentration of 2mM to detach the cells from the walls of the wells. Then transfer the cells to the tubes.

5.2.2 Use flavivirus (Zika virus and Dengue virus) to infect the incubated PBMCs

For flavivirus experiments, incubate about one million cells with Zika virus (strain MEX_CIENI551, MOI=10 and 5) and Dengue virus (clone BRAZIL, MOI=10 and 5) and mock in one well of 48-well plate. Incubate the cells for 48h at 37℃, 5% CO2. At 3h before the end of the incubation, add Brefeldin A to block the secretion of the cytokines. At the end of the incubation, add EDTA at the concentration of 2mM to detach the cells from the walls of the wells. Then transfer the cells to the tubes.

5.3 Cell fixation, permeablization and antibody staining (before PLAYR protocol)

Cells at a density of ~1 × 10^6/mL were fixed in RPMI medium without serum in 1.6% paraformaldehyde (Electronic Microscopy Sciences) for 10 min at room temperature under gentle agitation as described previously49. For detection of protein epitopes, cells were stained with antibodies in PBS (Life Technologies) supplemented with 5 mg/mL UltraPure BSA (Life Technologies), 0.2% saponin (Sigma-Aldrich), 2.5% vol/vol polyvinylsulfonic acid (Polysciences), and 40 U/mL RNasin (Promega) for 30 min at room temperature. After washing, antibodies were cross-linked to the cells with 5 mM bis(sulfosuccinimidyl) suberate (Pierce) in a buffer containing PBS, 0.2% saponin, and 40 U/mL RNasin for 30 min at room temperature at a density of ~20 × 10^6 cells/mL. Glycine was added to a final concentration of 100 mM, and samples were incubated for 5 min. Cells were pelleted and permeabilized with ice-cold methanol for at least 10 min on ice. Once in methanol, cells can be stored at –80 °C for several weeks without loss of antibody signal or RNA degradation. For detection of RNA only, cells were permeabilized in ice-cold methanol immediately after fixation with paraformaldehyde.

For mass cytometry, antibodies to the following were purchased: CD19 (Longwood, HIB19), CD4 (Longwood, RPA-T4), CD8 (Biolegend, RPA-T8), CD14 (DVS, M5E2), CD123 (Longwood, 6H6), CD45 (DVS, HI30), CD11c (DVS, Bu15), CD16 (DVS, 3G8), CD3 (Longwood, UCHT1), HLA-DR (Longwood, L243), CD56 (Longwood, NCAM16.2), IL-1b (H1b-27), CXCL-10 (IP-10), IL-8 (BioLegend, E8N1), IFNb (Longwood, IFNb/A1), TNFa (Longwood, MAb11), 4G2 (NOVUS, NBP2-52709).

5.4 PLAYR protocol

PLAYR probes were designed using the PLAYRDesign software developed in-house (available at https://github.com/nolanlab/PLAYRDesign). PLAYR probes were synthesized at the Yale Keck Biotechnology Resource Laboratory and resuspended in diethylpyrocarbonate (DEPC)-treated water at a concentration of 100 µM. The carrier solution for most of the protocol steps, including washes, was PBS, 0.1% Tween (Sigma-Aldrich) and 4 U/mL RNasin. Paraformaldehyde-fixed and methanol-permeabilized cells (see above) were pelleted by centrifugation at 600g for 3 min. Hybridizations with PLAYR probes were performed in a buffer based on DEPC-treated water (Life Technologies) containing 1× saline-sodium citrate (SSC) (Affymetrix), 2.5% vol/vol polyvinylsulfonic acid, 20 mM ribonucleoside vanadyl complex (New England BioLabs), 40 U/mL RNasin, 1% Tween, and 100 µg/mL salmon sperm DNA (Life Technologies). PLAYR probes for all target transcripts of an experiment were mixed and heated to 90 °C for 5 min. Probes were then chilled on ice and added to cells in hybridization buffer at a final concentration of 100 nM. Cells were incubated for 1 h at 40 °C under vigorous agitation and subsequently washed three times. Cells were then incubated for 20 min in a buffer containing PBS, 4× SSC and 40 U/mL RNasin at 40 °C under vigorous agitation. Samples to be analyzed by mass cytometry were barcoded at this step. After two washes, cells were incubated with 100 nM insert/backbone

oligonucleotides in PBS, 1× SSC and 40 U/mL RNasin for 30 min at 37 °C. After two washes, cells were incubated for 30 min with T4 DNA ligase (Thermo) at room temperature with gentle agitation and then for 16 h with phi29 DNA polymerase (Thermo) at 30 °C under agitation. Longer amplification (up to 16 h) generally increases signal intensity. Both enzymes were used according to their respective manufacturers' instructions, with the addition of 40 U/mL RNasin. For mass cytometry, cells were incubated with metal-conjugated detection oligonucleotides at a concentration of 10 nM for 30 min at 37 °C in PBS, 5 mg/mL BSA, 0.02% sodium azide. After washing, incubate the cells for 15 minutes at room temperature with 1× Intercalator Solution (Fluidigm). The Ir-Intercalator acts as a cell-identifying reagent during CyTOF analysis. The 1.6% PFA in the Intercalator Solution fixes the cell samples to prevent loss of metal and antibody when they are suspended in ddH2O at Step 16 before mass cytometry analysis. Intercalating cell samples may be stored for up to 1 month at 4 °C. Wash cells with washing buffer for once, with ddH2O for twice, and then analyze on a mass cytometer, collecting all pooled samples as a single FCS file.

5.5 Preparation of metal-conjugated detection oligonucleotides

Maleimide-activated Maxpar metal-chelating X8 polymers (Fluidigm, Maxpar labeling kit) were loaded with metals and purified using centrifugal filters as the manufacturer's instructions. Detection oligonucleotides carrying a 5' Thiol-Modifier C6 S-S (Glen Research) were synthesized at the Stanford Protein and Nucleic Acid Facility. Oligonucleotides were resuspended in DEPC-treated water at 250 µM, and the thiol was reduced by treatment with 50 mM TCEP (Pierce) for 30 min at room temperature. After ethanol precipitation, oligonucleotides were resuspended in C buffer (Fluidigm, Maxpar labeling kit), and conjugation reactions were performed with 2 nmol of

oligonucleotide per reaction with X8 polymer. After 2 h at room temperature, TCEP was added to a final concentration of 5 mM, and samples were incubated for 30 min to reduce unconjugated oligonucleotides. Conjugates were filtered through 30-kDa centrifugal filter units (EMD Millipore) in a total of 500 µl of water, spun at 14,000g for 12 min, and washed twice with DEPC-treated water (Life Technologies). Purified detection oligonucleotide conjugates were resuspended in DEPC-treated water at a concentration of 1 µM and stored at 4 °C. The metals were purchased from DVS: MaxPAR X8 Ab label Kit, 145Nd; MaxPAR X8 Ab label Kit, 148Nd; X8 Ab Label Kit, 158Gd; MaxPAR X8 Ab label Kit, 164Dy; DN3 AB labeling kit, 165Ho; MaxPAR X8 Ab label Kit, 171Yb; MaxPAR X8 Ab label Kit, 173Yb.

5.6 Debarcoding process

Barcoding kit was purchased from MaxPAR. The debarcoding software, a detailed manual covering its installation and usage, and a sample dataset is available at https://github.com/nolanlab/single-cell-debarcoder/releases/ latest. Modify sample names in the barcoding key CSV template file with text editing software or Microsoft Excel, and save modified version. Open the Single Cell Debarcoder and when prompted by the dialog box, select the correct saved barcode key. Click the ‘Select FCS File’ button in the ‘Select files’ panel and choose the FCS file containing the pooled samples. When the preliminary debarcoding is done, which may take several seconds depending on the size of the FCS file, a plot of the cell counts of the barcode populations will appear. To determine which barcode population corresponds to a particular line on the plot, click on that line and the label from the barcode key will appear. Enter a ‘separation threshold’ in the input box in the ‘Filters’ panel. This is a number between 0 and 1 that defines the minimum distance after normalization between the ‘positive’ and ‘negative’ barcode channels that is required for a cell to be assigned to a barcode

channel. The goal is to filter out uncertain barcode assignments but still retain sufficiently large barcode populations; a suggested starting point is a value just below which the cell count of the barcode populations dramatically decreases. A barcode well that has been left blank can provide a guidance as to where to set the threshold, and also provides an estimate of false assignment rate. In a well stained sample, the false assignment rate calculated from cells assigned to wells left blank is typically <0.5%. Valuate the barcode separation threshold by browsing through the 'Event' plots of different barcode populations. It may be useful to use the 'zoom' and 'pan'tools. Adjust the threshold if necessary, settling on the smallest value for which all the populations consist of cells in which positive and negative barcodes are sufficiently separated. Check the barcode deconvolution with biaxial plots of the barcode channels. To filter outliers from the populations, decrease the Mahalanobis threshold. This process of adjusting the separation threshold and the Mahalanobis threshold may be iterative until the appropriate balance between cell number and deconvolution confidence is reached. Examples of choosing the parameters can be found in the manual. Create a separate FCS file for each barcode population by pressing the 'Save Debarcoded Files' button, which will give a prompt to select the folder in which to save the files. By default the names of the wells as entered in the barcode key are appended to the original file name, but that may be adjusted prior to saving by editing the base file name next to the 'Save Debarcoded Files' button.

Chapter Six References

[0] 上海交通大学，《上海交通大学本科生毕业设计（论文）撰写规范》，12,6,2012

[1] Jaitin, D.A. et al. Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. Science 343, 776–779 (2014).

[2] Hashimshony, T., Wagner, F., Sher, N. & Yanai, I. CEL-Seq: single-cell RNA-Seq by multiplexed linear amplifcation. Cell Rep. 2, 666–673 (2012).

[3] Islam, S. et al. Highly multiplexed and strand-specifc single-cell RNA 5’ end sequencing. Nat. Protoc. 7, 813–828 (2012).

[4] Islam, S. et al. Quantitative single-cell RNA-seq with unique molecular identifers. Nat. Methods 11, 163–166 (2014).

[5] Ramsköld, D. et al. Full-length mRNA-seq from single-cell levels of RNA and individual circulating tumor cells. Nat. Biotechnol. 30, 777–782 (2012).

[6] Sasagawa, Y. et al. Quartz-Seq: a highly reproducible and sensitive single-cell RNA sequencing method, reveals non-genetic gene-expression heterogeneity. Genome Biol. 14, R31 (2013).

[7] Shalek, A.K. et al. Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. Nature 498, 236–240 (2013).

[8] Wu, A.R. et al. Quantitative assessment of single-cell RNA-sequencing methods. Nat. Methods 11, 41–46 (2014).

[9] Deng, Q., Ramsköld, D., Reinius, B. & Sandberg, R. Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells. Science 343, 193–196 (2014).

[10] Picelli, S. et al. Full-length RNA-seq from single cells using Smart-seq2. Nat. Protoc. 9, 171–181 (2014).

[11] Tang, F., Lao, K. & Surani, M.A. Development and applications of single cell transcriptome analysis. Nat. Methods 8, S6–S11 (2011).

[12] Fan, H.C., Fu, G.K. & Fodor, S.P.A. Expression profling. Combinatorial labeling of single cells for gene expression cytometry. Science 347, 1258367 (2015).

[13] Dvinge, H. et al. Sample processing obscures cancer-specifc alterations in leukemic transcriptomes. Proc. Natl. Acad. Sci. USA 111, 16802–16807 (2014).

[14] Grün, D. & van Oudenaarden, A. Design and analysis of single-cell sequencing experiments. Cell 163, 799–810 (2015).

[15] Bauman, J.G., Bayer, J.A. & van Dekken, H. Fluorescent in-situ hybridization to detect cellular RNA by flow cytometry and confocal microscopy. J. Microsc. 157, 73–81 (1990).

[16] Patterson, B.K. et al. Detection of HIV-1 DNA and messenger RNA in individual cells by PCR-driven in situ hybridization and flow cytometry. Science 260, 976–979 (1993).

[17] Belloc, F. & Durrieu, F. Detection of mRNA species by flow cytometry. Methods Cell Biol. 42, 59–69 (1994).

[18] Borzì, R.M. et al. A fluorescent in situ hybridization method in flow cytometry to detect HIV-1 specifc RNA. J. Immunol. Methods 193, 167–176 (1996).

[19] Lalli, E., Gibellini, D., Santi, S. & Facchini, A. In situ hybridization in suspension and flow cytometry as a tool for the study of gene expression. Anal. Biochem. 207, 298–303 (1992).

[20] Just, T., Burgwald, H. & Broe, M.K. Flow cytometric detection of EBV (EBER snRNA) using peptide nucleic acid probes. J. Virol. Methods 73, 163–174 (1998).

[21] Larsson, C., Grundberg, I., Söderberg, O. & Nilsson, M. In situ detection and genotyping of individual mRNA molecules. Nat. Methods 7, 395–397 (2010).

[22] Weibrecht, I. et al. In situ detection of individual mRNA molecules and protein complexes or post-translational modifcations using padlock probes combined with the in situ proximity ligation assay. Nat. Protoc. 8, 355–372 (2013).

[23] Player, A.N., Shen, L.P., Kenny, D., Antao, V.P. & Kolberg, J.A. Single-copy gene detection using branched DNA (bDNA) in situ hybridization. J. Histochem. Cytochem. 49, 603–612 (2001).

[24] Porichis, F. et al. High-throughput detection of miRNAs and gene-specifc mRNA at the single-cell level by flow cytometry. Nat. Commun. 5, 5641 (2014).

[25] Frei, A.P. et al. Highly multiplexed simultaneous detection of RNAs and proteins in single cells. Nat Methods. 13(3): 269-275 (2016).

[26] Irish, J.M. et al. Single cell profling of potentiated phospho-protein networks in cancer cells. Cell 118, 217–228 (2004).

[27]. Gaudillière, B. et al. Clinical recovery from surgery correlates with single cell immune signatures. Sci. Transl. Med. 6, 255ra131 (2014).

[28] Jaye, D.L., Bray, R.A., Gebel, H.M., Harris, W.A.C. & Waller, E.K. Translational applications of flow cytometry in clinical practice. J. Immunol. 188, 4715–4719 (2012).

[29] Kaleem, Z. et al. Flow cytometric analysis of acute leukemias. Diagnostic utility and critical analysis of data. Arch. Pathol. Lab. Med. 127, 42–48 (2003).

[30] Bendall, S.C. et al. Single-cell mass cytometry of differential immune and drug responses across a human hematopoietic continuum. Science 332, 687–696 (2011).

[31] Shi, P-Y (editor) (2012). Molecular Virology and Control of Flaviviruses. Caister Academic Press. ISBN 978-1-904455-92-9.

[32] McLean, B.J. et al. A novel insect-specific flavivirus replicates only in Aedes-derived cells and persists at high prevalence in wild Aedes vigilax populations in Sydney, Australia. Virology. 486: 272–283 (2015).

[33] Suthar M.S. et al. West Nile virus infection and immunity. Nat. Rev. Microbiology. 11: 115-128 (2013)

[34] Fredriksson, S. et al. Protein detection using proximity-dependent DNA ligation assays. Nat. Biotechnol. 20, 473–477 (2002).

[35] Söderberg, O. et al. Direct observation of individual endogenous protein complexes in situ by proximity ligation. Nat. Methods 3, 995–1000 (2006).

[36] Lizardi, P.M. et al. Mutation detection and single-molecule counting using isothermal rolling-circle amplifcation. Nat. Genet. 19, 225–232 (1998).

[37] Bandura, D.R. et al. Mass Cytometry: Technique for Real Time Single Cell Multitarget Immunoassay Based on Inductively Coupled Plasma Time-of-Flight Mass Spectrometry. Anal. Chem. 81, 6813 (2009).

[38] Lou, X. et al. Polymer-based elemental tags for sensitive bioassays. Chem. Int. Ed. Engl. 46, 6111 (2007).

[39] Gorden, K.B. et al. Synthetic TLR agonists reveal functional differences between human TLR7 and TLR8. J Immunol. 174(3):1259-68 (2005).

[40] Maecher H.T. et al. Standardizing immunophenotyping for the Human Immunology Project. Nat. Rev. Immunol. 12: 191–200 (2012).

[41] Davis, M. M. A prescription for human immunology. Immunity 29, 835–838 (2008).

[42] Giesen C. et al. Highly multiplexed imaging of tumor tissues with subcellular resolution by mass cytometry. Nat. Methods, 11: 417-422 (2014).

Chapter Seven Supplementary

7.1 Set up working environment on Farnam for PLAYRDesign

7.1.1 Notes

1. Personal Yale NetID account is not directly related to personal Farnam account. It is OK if one tries to log in with one NetID but uses other's Farnam account as long as he/she has the private key file and knows the keyphrase.
2. It has been confirmed that the Farnam Cluster has been equipped with R (required version), R studio, Blast+ and Primer3.
3. The whole procedure is designed for Windows users. Mac users and Linux users will need to make some changes according to Yale YCRC website.

7.1.2 Procedures

1. Apply for an account for Farnam on YCRC: https://research.computing.yale.edu/support/hpc/account-request.
2. If you are using Windows, please download MobaXterm from https://mobaxterm.mobatek.net/. Although it is recommended to have an installed version, you can just download a portable version to bypass Yale ITS.

(If you are using other operation system, you may use different softwares according to https://research.computing.yale.edu/support/hpc/user-guide.

3. Use MobaXterm to apply for SSH and operate according to the instructions on https://research.computing.yale.edu/support/hpc/user-guide/connect-windows.

Save the two keys (public and private) into a location other than "/bin" (or they will be deleted after each session ends).

4. Open up a session in the MobaXterm. Log in your account with your keyphrase. (it will not appear on the interphase, just go on to type in the key)

5. To have a node in the Farnam Cluster (or MobaXterm session), run:

srun --x11 --pty -p interactive bash

6. Creat a file as working directory, run:

mkdir WORKING DIRECTORY

7. Load softwares on the Farnam node (or MobaXterm session), run:

module load RStudio #mind caps lock

module load BLAST+ #mind caps lock

module load Primer3 #mind caps lock

8. Install Devtools:

Open an R session, run the following command and select a CRAN mirror when prompted:

install.packages("devtools")

9. Install Bioconductor packages:

In the R session, run the following commands:

source("http://bioconductor.org/biocLite.R")

biocLite(c("AnnotationDbi", "AnnotationFuncs", "BSgenome", "BSgenome.Hsapiens.UCSC.hg19", "Biostrings", "GenomicFeatures", "S4Vectors", "GenomicRanges", "IRanges", "org.Hs.eg.db"))

10. Install PLAYRDesign:

In the R session, run the following commands:

library(devtools)

install_github("nolanlab/PLAYRDesign")

11. Generate two databases: one contains repetitive sequences for the organism of interest (i.e. human), the other is a database of all the transcripts in the organism of interest (i.e. human).

a) The repetitive sequences can be downloaded from http://www.girinst.org/server/RepBase/index.php. Choose file RepBase23.03.fasta.tar.gz. It will need you to apply for an account, which will take just about one day. Unpack the file and upload “humrep.ref” and “simple.ref” to the working directory. To concatenate them to generate the repbase.fa file, on the Moba session run:

cat human.ref simple.ref > repbase.fa

b) The Human RefSeq RNA sequences can be downloaded by visiting the NCBI ftp server (ftp://ftp.ncbi.nlm.nih.gov/), navigating to the refseq -> H_sapiens -> H_sapiens -> RNA folder and selecting the rna.fa.gz file. Unpack the file. To filter this file to only contain NR and NM records, in R session run:

library(PLAYRDesign)

PLAYRDesign.filter_refseq_file("rna.fa", "rna_human_high_qual.fa")

c) Convert them to BLAST+ databases by running these commands on Moba session:

makeblastdb -in repbase.fa -dbtype nucl

makeblastdb -in rna_human_high_qual.fa -parse_seqids -dbtype nucl

12. Configuring PLAYRDesign

a) On the Moba session, run:

nano playrdesign_conf.txt

b) Exit the text (according to the instructions at the bottom). On the Moba session run:

which blastn

copy THE DIRECTORY1 appearing on the screen (use the right click)

To open the text file, on the Moba session run:

! nano

In the text file, type:

BLASTN_EXEC= THE DIRECTORY1 #The full path to the blastn executable

c) Exit the text (according to the instructions at the bottom). On the Moba session run:

mkdir DB

mv rna* DB

mv repba* DB

cd DB

recreate the original files in the WORKING DIRECTORY.

copy THE DIRECTORY2 appearing on the screen (use the right click)

To exit exit the DB file, on the Moba session run:

cd ../

To open the text file, on the Moba session run:

! nano

In the text file, type:

BLASTN_DB= THE DIRECTORY2/

The directory containing your BLAST database files. Do not forget "/".

d) Exit the text (according to the instructions at the bottom). On the Moba session run:

which primer3

copy THE DIRECTORY3 appearing on the screen (use the right click)

To open the text file, on the Moba session run:

! nano

In the text file, type:

PRIMER3_EXEC= THE DIRECTORY3 # The full path to the primer3_core executable

e) Exit the text (according to the instructions at the bottom). On the Moba session run:

which primer3_config

copy THE DIRECTORY4 appearing on the screen (use the right click)

To open the text file, on the Moba session run:

! nano

In the text file, type:

PRIMER3_CONFIG= THE DIRECTORY4/

The primer3_config directory that is found in the primer3 distribution. Do not forget "/".

f) Exit the text (according to the instructions at the bottom).

7.1.3 Usage

1. To have a node in the Farnam Cluster (or MobaXterm session), run:

srun --x11 --pty -p interactive bash

2. Enter working directory, run:

mkdir WORKING DIRECTORY

3. Load softwares on the Farnam node (or MobaXterm session), run:

module load RStudio #mind caps lock

module load BLAST+ #mind caps lock

module load Primer3 #mind caps lock

4. Run the RStudio, on the Moba session, run:

rstudio

5. Run the program, In the R session, run:

library(PLAYRDesign)

PLAYRDesign.run()

6. On the Firefox page, select the files as instructed.

7. Create the target transcript file:

a) Download RefSeq transcripts from the NCBI nucleotide database in the FASTA format, preferably choosing NR and NM records. Choosing the longest isoform of the transcript. You have to change the heading into:

>gi|NUMBER|ref|YOUR ACCESSION|

Only the 4th part really matters.

b) Upload the file to the WORKING DIRECTORY on the Moba session.

c) Select the file on the Firefox page.

8. Click “Start analysis”

9. Select the primer pairs:

a) The candidate probes are displayed as red rectangles at the bottom of the interface. Each pair is identified by a unique number on the rectangle. If you click on a primer both oligos in a pair will be selected. Selected oligos appear in the "Select oligos" box, and can be removed from there if desired.

b) It is also possible to generate a primer pair by selecting individual oligos from two different primer3 pairs. To do so ALT+Click on the first and then ALT+Click on the second (to clear the working selection ALT+Click on any blank region of the plot). When you combine oligos from two different pairs, the "Select oligos" box will display their unique ids (which are different from the ids you see in the plot, and can be visualized by hovering over an oligo) separated by a "_" character.

10. Generate the primer pairs:

Once you have selected the primer pairs use the "Select PLAYR system" dropdown to select an insert system and enter an id for the first oligo. Our standard is for the 5' oligo (on the transcript) of a pair to be the first one and to have an odd number. Hit the "Write oligos" button and a text file with the .playrdesign_out.txt extension will appear in your working directory (the same directory where your fasta files are located). The directory will also contain BLAST and Primer3 output files which are your free to delete after the analysis is completed.

7.2 All the oligos used in the study (Primers, Insert, Backbone, Detection probes)

The sequences of the primer pairs, insert, backbone, detection probes were all shown in the excel file in the supplementary materials.

Chapter Eight Acknowledgement

First and for most, I would like to thank my tutor, Professor Ruth R. Montgomery (Department of Internal Medicine at Yale University School of Medicine) for providing me with this precious opportunity to study in her lab. I really appreciate her trust that give me a chance to conduct a project on my own. I would also like to thank my lab colleagues, Xiaomei Wang, Helen Zhao, Megan Cahill, Anna Malawista for their great help with my project. X.W and A.M taught me the basic protocols and principles of mass cytometry. H.Z and M.C taught me the methods to analyze my data obtained from CyTOF. Thank Shelly Ren for operating mass cytometry each time I brought my sample. Thank Robert Bjornson for teaching me how to use LINUX operating system and helping me with the computing part of my project. Thank Professor Guangjie Chen for agreeing to be my tutor of my home university. Last by not least, I would like to extend my sincere gratefulness to Zhiyuan College and my parents for funding my research at Yale University.