LETTER ARTICLE

Discrepant mRNA and Protein Expression in Immune Cells

Jiawei Li1,3, Yi Zhang2,3, Cheng Yang1,3,4,* and Ruiming Rong1,3,*

1Department of Urology, Zhongshan Hospital, Fudan University, Shanghai 200032, China; 2Biomedical Research Center, Institute for Clinical Sciences, Zhongshan Hospital, Fudan University, Shanghai 200032, China; 3Shanghai Key Laboratory of Organ Transplantation, Shanghai, China; 4Zhangjiang Institute of Fudan University, Shanghai 201203, China

Abstract: With the development of single-cell mRNA sequencing (scRNA-seq), researchers have attempted to identify new methods for performing in-depth studies of immune cells. However, the discrepancies between the mRNA levels and the levels of surface proteins have confused many researchers. Here, we report a significant and interesting phenomenon in which the mRNA and protein expression levels were mismatched in immune cells. We concluded that scRNA-seq should be combined with other sequencing methods in single-cell studies (e.g., CIT-seq). The simultaneous assessment of both mRNA and protein expression will enhance the precision and credibility of the results.

Keywords: Single-cell mRNA sequencing, immune cell, mRNA, protein, myeloid-derived cells, protein expression.

1. INTRODUCTION

Single-cell mRNA sequencing (scRNA-seq) has been widely used and was investigated in 2015. The innovative Drop-seq strategy was first presented by McCarroll et al., allowing the rapid profiling of thousands of individual cells. With this novel form of biotechnology, cells can be precisely separated into nanoliter-sized aqueous droplets containing specific barcodes that are integrated with cellular mRNA and sequenced together. Given that cells of various origins may exhibit different characteristics, mRNA transcripts from thousands of individual cells can be simultaneously analyzed [1]. In 2017, Bielas and colleagues established a droplet-based system that enabled the quantification of mRNA from thousands of single cells in each sample with some modifications and could characterize the transcriptome of individual cells to further elucidate complex biological systems [2]. Subsequently, scRNA-seq was proposed as an effective tool for several areas of research, including cancer, autoimmune disease, sepsis, and transplantation. While researchers can only distinguish cell clusters according to the sequence and expression of known mRNAs that have been previously published, methods by which new cell types and subsets can be identified remain a challenge.

With the development of scRNA-seq, researchers have attempted to identify new methods for performing in-depth studies of immune cells. However, the discrepancies between the mRNA levels and the levels of surface proteins have confused many researchers. Moreover, the correlation between the protein and mRNA expression levels in individual cells is altered under different conditions, such as the steady state, long-term state changes, and short-term adaptation. These changes demonstrate the complexity of gene expression regulation, especially during dynamic transitions. The spatial and temporal variations of mRNAs, as well as the local availability of compounds for protein biosynthesis, are all essential factors that may strongly influence the relationship between protein levels and their coding transcripts [3]. However, there is no consistent explanation or particular reason for the mismatch between mRNA and surface protein expression in immune cells. Although many experiments have been designed to solve this problem, technological limitations continue to be an objective obstacle.

Myeloid-derived suppressor cells (MDSCs) are heterogeneous progenitors and immature myeloid cells; these immunosuppressive cells play an important role in the immune system [4]. Human MDSCs can be further divided into granulocytic MDSCs (G-MDSCs, LinHLA-DRCD33CD11bCD14CD15) and monocytic MDSCs (M-MDSCs, LinHLA-DRCD33CD11bCD14CD15) [5]. In our recent study, we unexpectedly found that the expression of some mRNAs is significantly lower than that of the cell surface proteins in myeloid-derived cells. To explore the expression of transcripts in peripheral MDSCs (pMDSCs), we sorted cells by fluorescence-activated cell sorting (FACS) and further analyzed them using 10x Genomics. As shown in Fig. (1A and B), the mRNA and protein expression levels of CD14 (an important marker of M-MDSCs) were similar. Unexpectedly, the surface expression of CD15 was high (Fig. 1A), whereas the CD15 mRNA (FUT4) expression was
Discrepant mRNA and Protein Expression in Immune Cells

In 2017, a brief communication in Nature Biotechnology presented a tool termed REAP-seq (V1 reagent of CITE-seq), which can be used to measure both the levels of gene and protein expression in a single cell using DNA-labeled antibodies and droplet microfluidics; this technology allowed the identification of unknown cell types [6] (Fig. 1C). Based on this technology and recommended by the official technical support from 10x Genomics, total-seq (CITE-seq) was used to enhance the analysis in our study. We found that the CD15 protein expression was high in pMDSCs and that the CD14 protein expression was similar to its mRNA expression (Fig. 1D). This phenomenon suggested that in some immune cells, the mRNA expression does not align with the surface marker expression. This phenomenon might be associated with scRNA-seq, which should not be ignored in the analysis of these immune cells.

Based on our findings, we sought to determine whether the same phenomenon exists in other immune cells. Support for scRNA-seq and CITE-seq of the peripheral blood mononuclear cells (PBMCs) was obtained from 10x Genomics. A single-cell gene expression dataset for analysis was downloaded from 10x Genomics at https://support.10xgenomics.com/single-cell-gene-expression/datasets. By reanalyzing the data, we found that this phenomenon was common in several peripheral blood immune cells. We outweighed the mRNA and cell surface marker expression, and the results are presented in Fig. (1F & G) and Table 1. Some cell surface markers were found to correlate with their homologous mRNAs (e.g., CD8a, CD11b, CD16, CD19, CD20, and CD25) (Fig. 1F) and may be potential markers for cell identification. However, there were discrepancies between the CD3, CD4, CD27, CD34, and CD80 mRNA levels and those of their related proteins (Fig. 1G). Therefore, utilizing CD3, CD27, CD34, and CD80 mRNA to distinguish cell clusters would cause a cluster deficiency since these mRNAs were expressed at lower levels than their proteins. Since the CD4 and CD69 mRNA expression was opposite to the corresponding protein expression, these parameters cannot be used to identify cell clusters, as this mismatch could cause incorrect results.

According to the results described above, we believe that scRNA-seq alone is insufficient for immune cell research. In many studies, SMART-seq and RNA-seq have been used to identify differentially expressed genes, which can be markers of different cell clusters [7]. However, these methods are unreliable for the identification of novel cell types. Thus, such techniques should be combined with other sequencing methods in single-cell studies (e.g., CITE-seq). Recently, an experiment introduced Ab-seq by BD company, which has the same basic theory as Total-seq and found the intuitive visualization of protein-transcript relationships on a single cell level [8]. In addition, a recent review of scRNA-seq technologies in inflammatory bowel disease also mentioned that CITE-seq could help researchers with increasing data quality [9]. Another work by Professor Fernandez reported
the first comprehensive paper about immune cell heterogeneity involved in human plaques as well as PBMCs using CITE-seq [10]. Thus, we would like to put forward this circumstance (mismatch between mRNA and protein expression) via our study to remind researchers dissect their analysis work, especially when it comes to cell identifications. In addition, to avoid the potential errors that occur in scRNA-seq analysis because of the mismatch phenomenon, we suggest researchers consider using scRNA-seq combined with RNA-seq or SMART-seq, or more efficiently, using CITE-seq, to increase the data quality.

CONCLUSION
In conclusion, the simultaneous assessment of both mRNA and protein expression will enhance the precision and credibility of the results.

Table 1. mRNA expression vs. protein expression of different cell markers by CITE-seq.

| Markers     | Expression between mRNA and Protein |
|-------------|-------------------------------------|
| CD3         | Lower                               |
| CD4         | Mismatch                            |
| CD8a        | Match                               |
| CD11b (ITGAM) | Match                           |
| CD14        | Match                               |
| CD15 (FUT4) | Lower                               |
| CD16 (FCGR3A) | Match                           |
| CD19        | Match                               |
| CD20        | Match                               |
| CD25 (IL2RA) | Match                           |
| CD27        | Lower                               |
| CD28        | Lower                               |
| CD34        | Lower                               |
| CD56 (NCAM1) | Lower                           |
| CD62L (SELL) | Lower                           |
| CD69        | Mismatch                            |
| CD80        | Lower                               |
| CD86        | Lower                               |
| CD127 (IL7R) | Lower                           |
| CD137 (TNFRSF9) | Lower                       |
| CD197 (CCR7) | Lower                           |
| CD274       | Lower                               |
| CD278 (ICOS) | Lower                           |
| CD335 (NCR1) | Lower                           |
| PD-1 (PDCD1) | Lower                           |
| TIGIT       | Lower                               |

LIST OF ABBREVIATIONS
scRNA-seq = Single Cell mRNA Sequencing
MDSCs = Myeloid-Derived Suppressor Cells
pMDSCs = Peripheral MDSCs
FACS = Fluorescence-Activated Cell Sorting
PMBC = Peripheral Blood Mononuclear Cells

AUTHORS’ CONTRIBUTIONS
Jiawei Li analyzed data, collected literature and drafted the manuscript. Cheng Yang and Ruiming Rong conceived the proposal, revised the manuscript and provided funding support. Yi Zhang revised the manuscript.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE
Not applicable.

HUMAN AND ANIMAL RIGHTS
No Animals/Humans were used for studies that are the basis of this research.

CONSENT FOR PUBLICATION
Not applicable.

AVAILABILITY OF DATA AND MATERIALS
The datasets generated and/or analyzed during the current study are available from the corresponding author on reasonable request.

FUNDING
This study was supported by the National Key R&D Program of China (2018YFA0107501 to R.R), National Natural Science Foundation of China (81770747 and 81970646 to R.R).

CONFLICT OF INTEREST
The authors declare no conflict of interest, financial or otherwise.

ACKNOWLEDGEMENTS
Declared none.

REFERENCES
[1] Macosko, E.Z.; Basu, A.; Satija, R.; Nemesh, J.; Shekhar, K.; Goldman, M.; Tiross, I.; Bialas, A.R.; Kamitaki, N.; Martersteck, E.M.; Trombetta, J.J.; Weitz, D.A.; Sanes, J.R.; Shalek, A.K.; Regev, A.; McCarroll, S.A. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. Cell, 2015, 161(5), 1202-1214. http://dx.doi.org/10.1016/j.cell.2015.05.002 PMID: 26000488
[2] Zheng, G.X.; Terry, J.M.; Belgrader, P.; Ryvkin, P.; Bent, Z.W.; Wilson, R.; Ziraldo, S.B.; Wheeler, T.D.; McDermott, G.P.; Zhu, J.; Gregory, M.T.; Shuga, J.; Montesclaros, L.; Underwood, I.G.; Masquelier, D.A.; Nishimura, S.Y.; Schnall-Levin, M.; Wyatt, P.W.; Hindson, C.M.; Bharadwaj, R.; Wong, A.; Ness, K.D.; Beppu, L.W.; Deeg, H.J.; McFarland, C.; Loeb, K.R.; Valente, W.J.; Ericson, N.G.; Stevens, E.A.; Radich, J.P.; Mikkelsen, T.S.; Hindson, B.J.; Bielsa, J.H. Massively parallel digital transcriptional profiling of single cells. Nat. Commun. 2017, 8, 14049. http://dx.doi.org/10.1038/ncomms14049 PMID: 28091601
[3] Liu, Y.; Beyer, A.; Aebersold, R. On the Dependency of Cellular Protein Levels on mRNA Abundance. *Cell*, **2016**, *165*(3), 535-550. http://dx.doi.org/10.1016/j.cell.2016.03.014 PMID: 27104977

[4] Gabrilovich, D.I. Myeloid-derived suppressor cells. *Cancer Immunol. Res.*, **2017**, *5*(1), 3-8. http://dx.doi.org/10.1158/2326-6066.CIR-16-0297 PMID: 28052991

[5] Veglia, F.; Perего, M.; Gabrilovich, D. Myeloid-derived suppressor cells coming of age. *Nat. Immunol.*, **2018**, *19*(2), 108-119. http://dx.doi.org/10.1038/s41590-017-0022-x PMID: 29348500

[6] Peterson, V.M.; Zhang, K.X.; Kumar, N.; Wong, J.; Li, L.; Wilson, D.C.; Moore, R.; McClanahan, T.K.; Sadekova, S.; Klappenbach, J.A. Multiplexed quantification of proteins and transcripts in single cells. *Nat. Biotechnol.*, **2017**, *35*(10), 936-939. http://dx.doi.org/10.1038/nbt.3973 PMID: 28854175

[7] Ziegenhain, C; Vieth, B; Parekh, S; Reinius, B; Guillame-Ladoux, A; Smets, M. Comparative analysis of single-cell RNA sequencing methods. *Mol. Cell.*, **2017**, *65*(4), 631-643.e4.

[8] Mair, F.; Erickson, J.R.; Voillet, V.; Simoni, Y.; Bi, T.; Tyznik, A.J.; Martin, J.; Gottardo, R.; Newell, E.W.; Pricic, M. A targeted multi-omic analysis approach measures protein expression and low-abundance transcripts on the single-cell level. *Cell Rep.*, **2020**, *31*(1), 107499. http://dx.doi.org/10.1016/j.celrep.2020.03.063 PMID: 32268080

[9] Corridoni, D.; Chapman, T.; Antanaviciute, A.; Satsangi, J.; Simmons, A. Inflammatory bowel disease through the lens of single-cell rna-seq technologies. *Inflamm. Bowel Dis.*, **2020**, *26*(11), 1658-1668. http://dx.doi.org/10.1093/ibd/izaa081 PMID: 32386055

[10] Saigusa, R.; Ley, K. CITE-Seq hits vascular medicine. *Clin. Chem.*, **2020**, *66*(6), 751-753. http://dx.doi.org/10.1093/clinchem/hva08 PMID: 32091084