IMCell$^{\text{XMBD}}$: A statistical approach for robust cell identification and quantification from imaging mass cytometry images

Xu Xiao[1,2,#], Naifei Su[3,#], Yan Kong[4], Lei Zhang[5], Xin Ding[6], Wenxian Yang[3,*], Rongshan Yu[1,2,3,*]

[1] Department of Computer Science, School of Informatics, Xiamen University, Xiamen, China

[2] National Institute for Data Science in Health and Medicine, Xiamen University, Xiamen, China

[3] Aginome Scientific, Xiamen, China

[4] Peking University Cancer Hospital and Institute, Beijing, China

[5] School of Life Science, Xiamen University, Xiamen, China

[6] Zhongshan Hospital, Xiamen University, Xiamen, China

[#] These authors have contributed equally to this work.

[*] Corresponding author: rsyu@xmu.edu.cn

Imaging Mass Cytometry (IMC) has become a useful tool in biomedical research due to its capability to measure over 100 markers simultaneously. Unfortunately, some protein channels in IMC images can be very noisy, which may significantly affect the phenotyping results without proper data processing. We developed IMCell$^{\text{XMBD}}$[1], a highly effective and generalizable cell identification and quantification method for IMC images. IMCell performs denoising

[1] XMBD: Xiamen Big Data, a biomedical open software initiative in the National Institute for Data Science in Health and Medicine, Xiamen University, China.

by subtracting an estimated background noise value from pixel values for each individual protein channel, identifies positive cells from negative cells by comparing the distribution between segmented cells and decoy cells, and normalize the protein expression levels of the identified positive cells for downstream data analysis. Experimental results demonstrate that our method significantly improves the reliability of cell phenotyping which is essential for using IMC in biomedical studies.

1 Introduction

Analysis of the heterogeneity of cells is critical to discover the complexity and factuality of the life system. Recently, single-cell sequencing technologies have been increasingly used in the research of developmental physiology and disease [1–4], but the spatial context of individual cells in the tissue is lost due to tissue dissociation in these technologies. On the other hand, traditional immunohistochemistry (IHC) and immunofluorescence (IF) preserve spatial context but the number of markers is limited. The development of multiplex IHC/IF (mIHC/mIF) technologies, such as cyclic IHC/IF and metal-based multiplex imaging technologies [5–8], has enabled the detection of multiple markers simultaneously while preserving their spatial information. Imaging mass cytometry (IMC) [6,9], one of the metal-based mIHC technologies, uses a high-resolution laser with a mass cytometer and enables simultaneous measurement of up to 100 markers. Due to its high resolution and large number of concurrent marker channels available, IMC has been proven to be highly effective in identifying complex cell phenotypes and cell-cell interactions coupled with spatial locations, and has been utilized in many biomedical and clinical studies on tumor or immune diseases [6,10–21].

A number of methodological challenges must be overcome when applying IMC to clinical applications in order to derive reliable cell quantification and phenotyping results from IMC. Images generated by a mass cytometry system are subject to noise and other acquisition artifacts resulting from, e.g., sample protein degradation or signal spill-over between heavy metals [22]. Instrument performance can vary within a single sample, not to mention the technical variance among different instruments. Besides, the antibody performance and antigen retrieval condition can differ between samples due to their storage time and environment, which result in protein variations between and within samples. Therefore, specific data processing steps are needed to ensure measurement of cellular markers with high resolution, quality, and reproducibility. Quality control and data normalization have been incorporated into the standard operation procedures in the software of the mass cytometers to convert raw signals to images [23]. Most IMC image quality control and preprocessing steps are performed semi-automatically and tuned for individual datasets. Some generic signal processing techniques have been applied to different datasets, including background removal, hot pixels removal, and denoising by low pass filtering, etc. [11,24]. Data normalization has also been discussed to eliminate the variation between samples [25,26]. Despite the progress in IMC data processing tools, in practice it is still possible to obtain IMC images with very poor signal-to-noise ratios (SNR) that exceed the processing capabilities of existing tools. In such cases, it remains as an intricate issue to identify true positive cells from strong background noise and harmonize their protein expression levels across slides from different samples or different regions of interest (ROIs) from the same slide for downstream analysis.

In this paper, we present IMCell, a method for protein expression quantification for single cells from IMC images. IMCell is able to reliably identify positive cells from highly noisy channels of an IMC image, and perform expression quantification for these cells. To this end, IMCell uses a Monte Carlo method to create decoy cells randomly on the potential noise regions of the image, and computes the distribution of the protein expression of the decoy cells to derive the background noise level of the image. The positive cells are then identified with false discovery rate (FDR) control by comparing the protein expression distribution of decoy cell with that of the segmented true cells. To reduce the effect of background noise on the quantification results, IMCell further performs noise reduction on the IMC images with the identified background noise level. Finally, the protein expression values of the positive cells are normalized to mitigate the variations of pixel values across different IMC images. Our evaluation results show that IMCell can retain real signals with a user-defined confidence level and eliminate sample variations, improves IMC image quality, and benefits the downstream analysis.

2 Results

IMCell identifies true positive cells from noise IMCell identifies positive cells on each protein channel based on FDR control with the distribution of permuted decoy cells. First, IMCell randomly generates a large number of decoy cells on potential noise regions of each protein channel (Methods, Figure 1). With the generated decoy cells, IMCell identifies positive cells by comparing the distributions of cell protein expressions, calculated as the mean of pixel values of the cell, of all segmented cells and decoy cells, from which the detection threshold can be set based on the target FDR (Methods, Figure 1). Once the positive cells are identified on each protein channel, IMCell further estimates the background noise level (Methods), which is then removed from the respective IMC channel to generate a clean image for each channel.

We compared the performance for background noise removal of IMCell with two commonly used methods, the percentile method and the median filter. The percentile method defines a lower threshold T_l and an upper threshold T_h. It then removes outliers by setting pixel value to zero for those lower than T_l, and setting pixel values to T_h for those higher than T_h. Here we used the 1st percetile (Q_1) as T_l and the 99th percentile ($Q_9 9$) as T_h. Results show that the percentile method removes outliers but cannot deal with noise of similar intensity values as the signal, such as salt-and-pepper noise. On the other hand, the median filter is only effective in removing salt-and-pepper noise but does not remove other types of noise. In addition, it tends to remove true expression signals wrongly at cell boundaries, or if the true expression signals have a salt-and-pepper noise-like spatial patterns. In contrast, by estimating the background noise level from decoy cells randomly drawn from the potential noise regions of the image, IMCell successfully removed background noise while preserving the true protein expression values from positive cells, resulting in a cleaner image with significantly improved the SNR (Figure 2a, 2b).

We further compared the co-expression patterns of CD45, CD3 and CD4 from different methods and observed that IMCell can retain true CD3 signal since most CD4 T cells expressed both CD3 and CD4, while the median filter over-removed CD3 signal and the percentile method failed to remove noise in the CD3 channel (Figure 2c, 2d)).

IMCell reduces variations in pixel intensity and cell protein expression across IMC images. Analysis of the raw images and segmented cells show that the range of pixel intensity values and the level of SNR vary significantly among samples (Figure 3a). The difference is conspicuous even after performing the variance stabilizing transform, e.g., the inverse sinh transform [27], on the IMC images to reduce the overall range of the pixel intensities (Figure 3b). The distribution plots demonstrate that the variation across samples exists not only at pixel level but also at cell level, if the cell protein expressions were calculated directly from the raw images. Large inter-sample distribution variation could be misleading in downstream data analysis, as the cells may cluster by samples but not by cell types. In IMCell, protein expression levels are normalized across the entire dataset based on the identified positive cells (Methods). Figure 3c shows the variation of intensity across three samples at both pixel and cell levels after intensity normalization by IMCell.

IMCell enables clustering with biological significance To investigate the effects of different IMC image preprocessing methods on downstream analysis, we applied unsupervised clustering on cells generated from raw IMC images, images processed with the median filter, the percentile method, and IMCell, respectively, using a same subset of proteins as features. After clustering, the cell type of each cluster can be identified based on its marker expression pattern compared to that of known immune and tumor cell types (Figure 4). The cell types of the cell clusters obtained from raw IMC images or images processed using the percentile method can hardly be identified. As the heatmap shows, some clusters have more than one relatively high cell-type-specific protein expressions (Figure 4a). For example, Cluster 1 from the raw IMC images contains similar protein expression level for both lymphoid (CD4) and myeloid cells (CD14, CD68), causing confusion in cell type identification. The percentile method also leads to a confusing heatmap where the cell types cannot be ascertained (Figure 4b). Alternatively, by applying the median filter or IMCell on the raw images, the cell clustering results are more biologically significant (Figure 4c, 4d). For the clustering results obtained from cells of IMC images preprocessed by the median filter, we can annotate Cluster 12 as B cell, but still have difficulty to determine other two clusters (Cluster 1 and 10) because they contain T cell markers (e.g., CD4, CD8) and a certain amount of myeloid cell markers such as CD68 and CD14. On the other hand, we are able to obtain highly specific cell clusters from clustering results obtained from cells quantified with IMCell, e.g., CD4 T cell (Cluster 4), CD8 T cell (Cluster 1), B cell (Cluster 3) and myeloid cell (Cluster 12, 13, 15).

3 Discussion

In this work, we developed IMCell which enables efficient and accurate cell quantification from IMC images. Our work is based on statistical testing on the distributions of both segmented cells, which are regarded as true cells identified by image segmentation software, and decoy cells. As decoy cells are drawn from potential noise-only regions of IMC image with random shapes and locations, it can be anticipated that its distributions will highly resemble those of negative cells (i.e., cells that don’t express target proteins). Therefore, the positive cells can be reliably identified with proper FDR control base on the distributions of both cells. Note that the successful application of IMCell depends on the availability of information on true cell segmentation. In this work we used Dice-XMBD [28], a deep neural network based IMC cell segmentation tool that is able to perform automatic cell segmentation from IMC images without manual annotation. It is also possible to use other cell segmentation tools, e.g., Ilastik [29] and CellProfiler [30], to perform such a task.

Normalization across different images is critical to align the protein expressions to the same sea-level such that they can be compared in downstream data analysis. However, such normalization can only be performed if the positive cells (i.e., cells expressing certain target proteins) can be reliably identified. Otherwise, the normalization can falsely amplify negative cells located at noise regions of the image, resulting in severe false positive issues that plague the downstream biological analysis. For this reason, expression normalization is seldom performed in existing IMC processing pipelines although significant inter-slide variations of marker protein expressions are common in IMC studies. In IMCell, by rigorous FDR control, expression normalization is only performed on high-confidence positive cells, thus minimizing the risk of amplification of false-positive cells. As validated by visual inspection and clustering analysis, cell quantification by IMCell leads to much more consistent connections between cell phenotypes and marker protein expressions. We anticipate that IMCell could help to promote better usage of the IMC technologies both in research labs and in clinical settings.

4 Methods

Patients and IMC data acquisition Melanoma cancer formalin-fixed paraffin-embedded (FFPE) tissues were stained with a customized panel (35 antibodies) to generate the IMC images used in this study. We excluded images containing large areas with nonspecific background staining that could be caused by nonspecific antibody binding [31] by manual inspection using the MCD viewer

(V1.0.560.6). The remaining 158 images were further analyzed in the following procedures.

Overview of the IMCell workflow. IMCell consists of two main modules, denoising and normalization (Figure 5). Firstly, raw IMC images are preprocessed and segmented by any cell segmentation method. Then we randomly generate a number of decoy cells on the potential noise region of each protein channel image. The protein expressions of the decoy cells are used to estimate the background noise of the protein image. After that the protein expression distributions of all segmented cells and decoy cells are compared to identify positive cells with FDR control. Next, in the normalization module, to fairly compare positive cells across images, we scale the mean expression of positive cells from each image to the same level. More details are described as follows step by step.

Cell segmentation using Dice-XMBD Single cells were identified by Dice-XMBD [28] using a pretrained deep neural network model, and referred to as segmented cells in this paper. Note that other cell segmentation methods can also be used in the IMCell workflow. For quality control, the segmented cells that cover less than 5 pixels are discarded. The cell protein expressions are extracted as the mean of the pixel intensity values in each cell mask region.

Preprocessing and hot pixel removal We first applied the hyperbolic inverse sine function (arcsinh) on all the pixel intensities for each channel. The raw marker intensities output from cytometers tend to have strongly skewed distributions with varying ranges of expression values. It is thus a common practice to transform the raw marker intensities using arcsinh to make the distributions more symmetric and to map them to a comparable range of expressions [27,32].

Hot pixels were removed by filtering with a 5×5 pixel^2 window. If the center pixel of the window was in the top 2% of all pixel intensity values in the channel and was at least $4\times$ above the median value of all pixels in the window, it will be identified as a hot pixel and its value will be replaced by the median value in the window. This step reduces the scattered hot pixels' noise on quantification of protein expression values for the cells.

Generating decoy cells We established the distribution of noise for each channel by generating a large number (N) of decoy cells using a Monte Carlo method. To this end, we first identified regions on the image that potentially contain noise-only signals without real protein expression by excluding pixels with values above $0.05 \times Q_{99}$, where Q_{99} is the 99th percentile (Q_{99}) of the pixel intensity values. After that, we set the value of remaining pixels to zero and smooth the noise regions by applying a 5×5 median filter on the image.

We then fit each segmented cell as an ellipse. For each image, the mean and variance of the major axis, the minor axis, and the orientation angle of all the segmented cells were calculated, and these three parameters were fit using individual Gaussian models. Random parameters are drawn from the distributions of the major axis, the minor axis, and the orientation, respectively, to form an ellipse as a decoy cell. The decoy cell was randomly placed in the noise region of the channel image, such that the center of the decoy cell was at least 5 pixels away from image boundaries. The decoy cell should only lie in noise regions, i.e., all of its pixels lie in noise regions as in the noise region mask. When the decoy cell lies on the border of the image, it must cover more than 5 pixels in the image, otherwise it will be discarded. We further filter out the decoy cell if the area it covered exceeded the size range of all segmented cells. Then, the protein expression value for each decoy cell was calculated as the mean of its pixel intensities in the preprocessed IMC image.

Background noise removal To eliminate the effect of different background noise profiles and levels between different proteins in an IMC dataset, we removed background noise using the decoy cells generated from the noise regions. For each protein channel, the mean of protein expressions of all generated decoy cells was calculated, which was further subtracted from each pixel intensity to remove channel-specific background noise.

Positive cells identification by FDR control Note that the segmented cells may include both positive cells and negative cells. We used a permutation test to compare the protein expression distributions between segmented cells and randomly drawn decoy cells from the noise regions, and use FDR control to identify positive cells. The FDR value can be adjusted to obtain positive cells with acceptable error-tolerant rate. The FDR of true cell identification is calculated by

$$FDR = \frac{FP}{FP + TP}, \tag{1}$$

where TP and FP refer to true positive and false positive, respectively. More specifically, TP refers to the number of segmented cells with protein expression values larger than the threshold, while FP refers to the number of decoy cells with protein expression values larger than the threshold. The default value of FDR was set to 0.01, and the threshold for positive cell identification can be then determined to satisfy the FDR level.

Normalization of cell protein expressions The data processing steps above are all performed on individual protein channel images. As the antibody performance and the SNR can differ considerably between FFPE tissues due to variations in tissue processing, we further normalized the cell protein expression values across different samples within one IMC dataset for each protein separately. Denote the channel image of protein p for sample i as $I_i^{(p)}$ and the mean of the protein expression values for all identified positive cells as $\mu_i^{(p)}$. Let $m^{(p)}$ denote the maximum protein expression value among all identified positive cells for protein p in all samples. The cell protein expression values for sample i were then scaled by factor $\frac{m^{(p)}}{\mu_i^{(p)}}$.

Single cell clustering and phenotyping High-dimensional single cell protein expression data were clipped at the 99th percentile followed by min-max normalization. We selected 20 markers to perform cell clustering: CD45, CD3, CD4, CD8a, FoxP3, CD20, CD68, CD14, CD16, CD11c, CD11b, IDO, Vimentin, α-SMA, E-cadherin, EpCAM, CA9, VEGF, PDGFRb and Collagen. The clustering analysis consists of two consecutive steps, first, a self-organizing map (50×50 nodes) implemented in FlowSOM (R package, v1.18.0) was used to generate several groups, then a community detection algorithm by Phenograph (R package, v0.99.1) was used on the mean expression values of each group from FlowSOM clustering results. Cell phenotyping was determined by calculating the mean of protein expressions for each cluster and compare the protein expression patterns of each cluster with that of known cell types.

Conflict of Interest Statement

RY and WY are shareholders of Aginome Scientific. The authors declare no other conflict of interest.

Author Contributions

WY and RY supervised the study and developed the concept. NS implemented the denoising method and XX conducted experiments in evaluation and biological analysis. YK and LZ worked on samples from patients. XD provided confirmatory pathology analyses. All authors wrote and discussed on the manuscript.

References

1. Lähnemann, D. *et al.* Eleven grand challenges in single-cell data science. *Genome Biology* **21**, 1–35 (2020).

2. Stubbington, M. J., Rozenblatt-Rosen, O., Regev, A. & Teichmann, S. A. Single-cell transcriptomics to explore the immune system in health and disease. *Science* **358**, 58–63 (2017).

3. Potter, S. S. Single-cell rna sequencing for the study of development, physiology and disease. *Nature Reviews Nephrology* **14**, 479–492 (2018).

4. Papalexi, E. & Satija, R. Single-cell rna sequencing to explore immune cell heterogeneity. *Nature Reviews Immunology* **18**, 35 (2018).

5. Tan, W. C. C. *et al.* Overview of multiplex immunohistochemistry/immunofluorescence techniques in the era of cancer immunotherapy. *Cancer Communications* **40**, 135–153 (2020).

6. Giesen, C. *et al.* Highly multiplexed imaging of tumor tissues with subcellular resolution by mass cytometry. *Nature Methods* **11**, 417–422 (2014).

7. Zrazhevskiy, P. & Gao, X. Quantum dot imaging platform for single-cell molecular profiling. *Nature Communications* **4**, 1–12 (2013).

8. Angelo, M. *et al.* Multiplexed ion beam imaging of human breast tumors. *Nature Medicine* **20**, 436–442 (2014).

9. Chang, Q. *et al.* Imaging mass cytometry. *Cytometry Part A* **91**, 160–169 (2017).

10. Damond, N. *et al.* A map of human type 1 diabetes progression by imaging mass cytometry. *Cell Metabolism* **29**, 755–768 (2019).

11. Wang, Y. J. *et al.* Multiplexed in situ imaging mass cytometry analysis of the human endocrine pancreas and immune system in type 1 diabetes. *Cell Metabolism* **29**, 769–783 (2019).

12. Ramaglia, V. *et al.* Multiplexed imaging of immune cells in staged multiple sclerosis lesions by mass cytometry. *Elife* **8**, e48051 (2019).

13. Böttcher, C. *et al.* Single-cell mass cytometry reveals complex myeloid cell composition in active lesions of progressive multiple sclerosis. *Acta Neuropathologica Communications* **8**, 1–18 (2020).

14. de Vries, N. L., Mahfouz, A., Koning, F. & de Miranda, N. F. Unraveling the complexity of the cancer microenvironment with multidimensional genomic and cytometric technologies. *Frontiers in Oncology* **10**, 1254 (2020).

15. Brähler, S. *et al.* Opposing roles of dendritic cell subsets in experimental gn. *Journal of the American Society of Nephrology* **29**, 138–154 (2018).

16. Aoki, T. *et al.* Single-cell transcriptome analysis reveals disease-defining t-cell subsets in the tumor microenvironment of classic hodgkin lymphoma. *Cancer Discovery* **10**, 406–421 (2020).

17. Jackson, H. W. *et al.* The single-cell pathology landscape of breast cancer. *Nature* **578**, 615–620 (2020).

18. Ali, H. R. *et al.* Imaging mass cytometry and multiplatform genomics define the phenogenomic landscape of breast cancer. *Nature Cancer* **1**, 163–175 (2020).

19. Dey, P. *et al.* Oncogenic kras-driven metabolic reprogramming in pancreatic cancer cells utilizes cytokines from the tumor microenvironment. *Cancer Discovery* **10**, 608–625 (2020).

20. Zhang, Y., Gao, Y., Qiao, L., Wang, W. & Chen, D. Inflammatory response cells during acute respiratory distress syndrome in patients with coronavirus disease 2019 (covid-19). *Annals of Internal Medicine* (2020).

21. Schwabenland, M. *et al.* Deep spatial profiling of human covid-19 brains reveals neuroinflammation with distinct microanatomical microglia-t-cell interactions. *Immunity* **54**, 1594–1610.e11 (2021).

22. Chevrier, S. *et al.* Compensation of signal spillover in suspension and imaging mass cytometry. *Cell Systems* **6**, 612–620 (2018).

23. Lee, B. H. & Rahman, A. H. Acquisition, processing, and quality control of mass cytometry data. In *Mass Cytometry*, 13–31 (Springer, 2019).

24. Baharlou, H., Canete, N. P., Cunningham, A. L., Harman, A. N. & Patrick, E. Mass cytometry imaging for the study of human diseases—applications and data analysis strategies. *Frontiers in Immunology* **10**, 2657 (2019).

25. Ijsselsteijn, M. E., Somarakis, A., Lelieveldt, B. P., Hollt, T. & de Miranda, N. F. Semi-automated background removal limits loss of data and normalises the images for downstream analysis of imaging mass cytometry data. *bioRxiv* (2020).

26. Keren, L. *et al.* A structured tumor-immune microenvironment in triple negative breast cancer revealed by multiplexed ion beam imaging. *Cell* **174**, 1373–1387 (2018).

27. 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).

28. Xiao, X. *et al.* Dice-XMBD: Deep learning-based cell segmentation for imaging mass cytometry. *Frontiers in Genetics* **12**, 1532 (2021).

29. Sommer, C., Straehle, C., Köthe, U. & Hamprecht, F. A. Ilastik: Interactive learning and segmentation toolkit. In *2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, 230–233 (2011).

30. Jones, T. R. *et al.* CellProfiler Analyst: data exploration and analysis software for complex image-based screens. *BMC Bioinformatics* **9**, 482 (2008).

31. Buchwalow, I., Samoilova, V., Boecker, W. & Tiemann, M. Non-specific binding of antibodies in immunohistochemistry: fallacies and facts. *Scientific Reports* **1**, 1–6 (2011).

32. Bruggner, R. V., Bodenmiller, B., Dill, D. L., Tibshirani, R. J. & Nolan, G. P. Automated identification of stratifying signatures in cellular subpopulations. *Proceedings of the National Academy of Sciences* **111**, E2770–E2777 (2014).

Figure captions

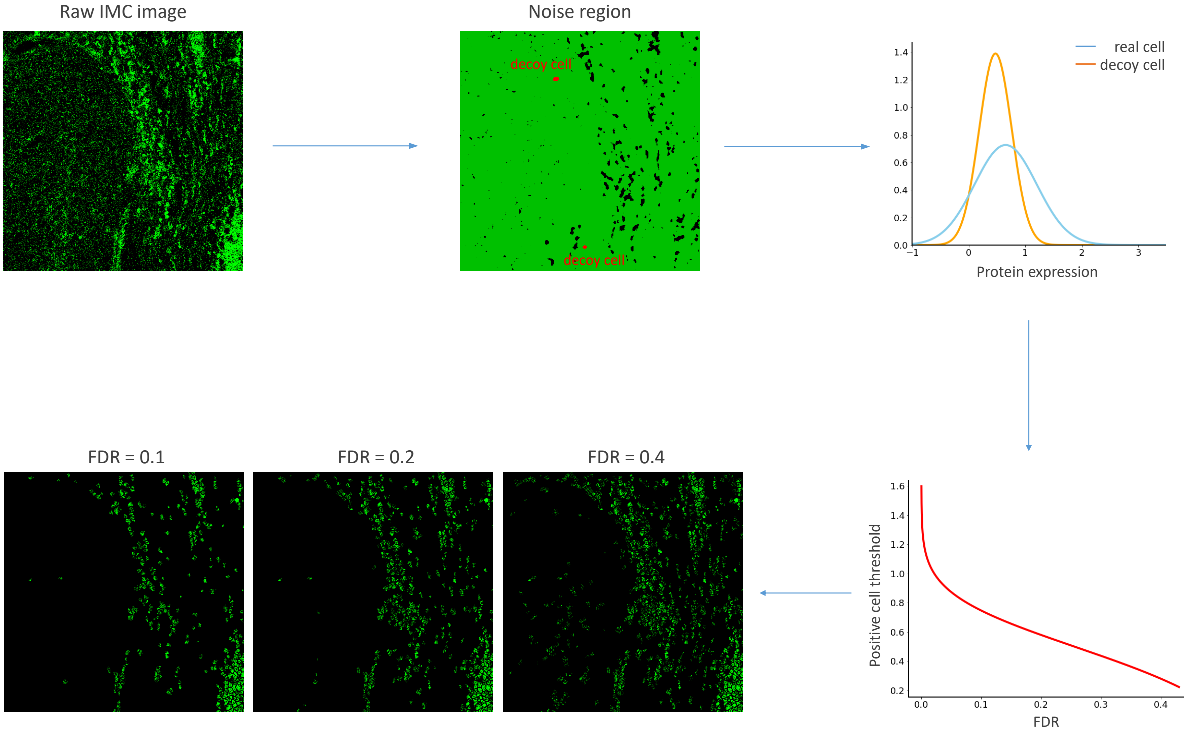


Figure 1: Positive cells identified by IMCell with different FDR control (sample: 76 ROI18, protein: CD74).

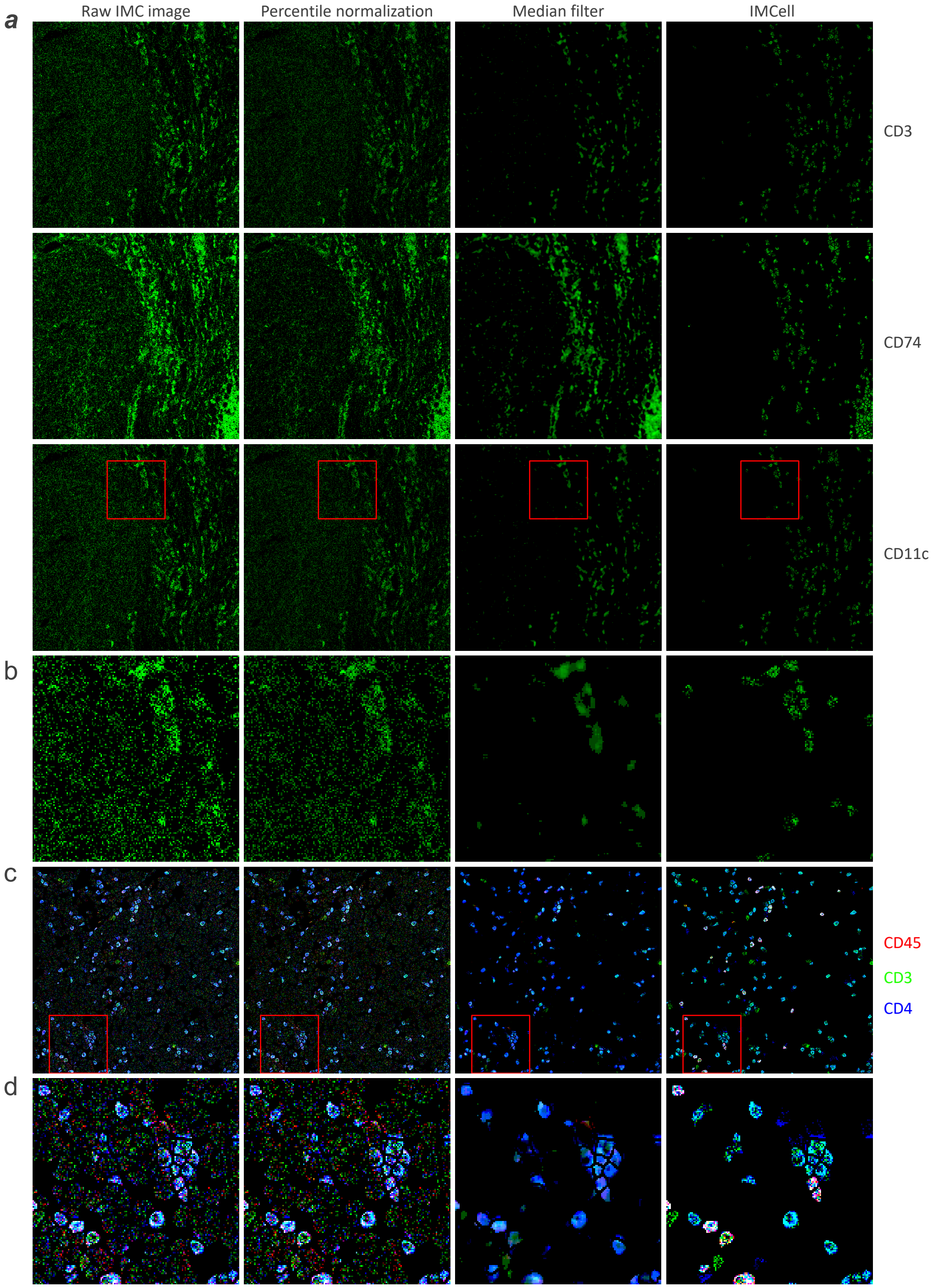


Figure 2: Performance evaluation of different methods. (a) Comparisons of cell identification results from raw IMC images, with 1^{st}-99^{th} percentile method to remove outliers, with median filter to remove salt-and-pepper noise, and with IMCell (sample: 76 ROI18). The red box marks the zoomed in areas on the below side (b) depicting the CD11c marker. Expression pattern of multi-markers (CD45, CD3, and CD4) in the whole images (c) and zoom-in areas (d).

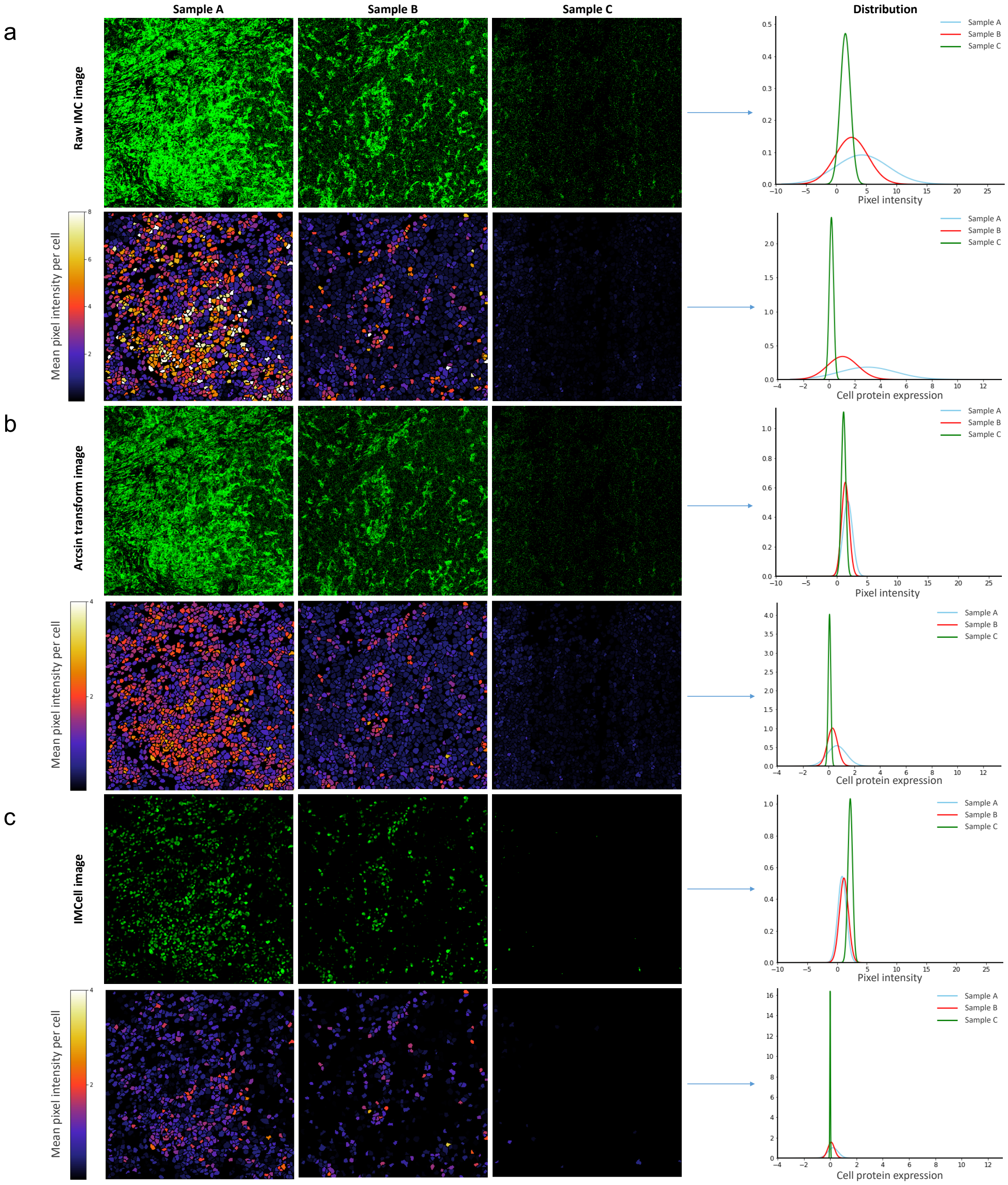


Figure 3: Variation of pixel intensity and cell protein expression across three samples (sample A: 65 ROI13, sample B: 65 ROI18, sample C: 33 ROI11). The left column shows (a) the pixel intensity (first row) and cell protein expression (second row) from the raw images, (b) the pixel intensity (first row) and cell protein expression (second row) from arcsinh-transformed images, and (c) the pixel intensity (first row) and cell protein expression (second row) from images processed by IMCell. The right column plots the distribution of the corresponding value (i.e., pixel intensity and cell protein expression).

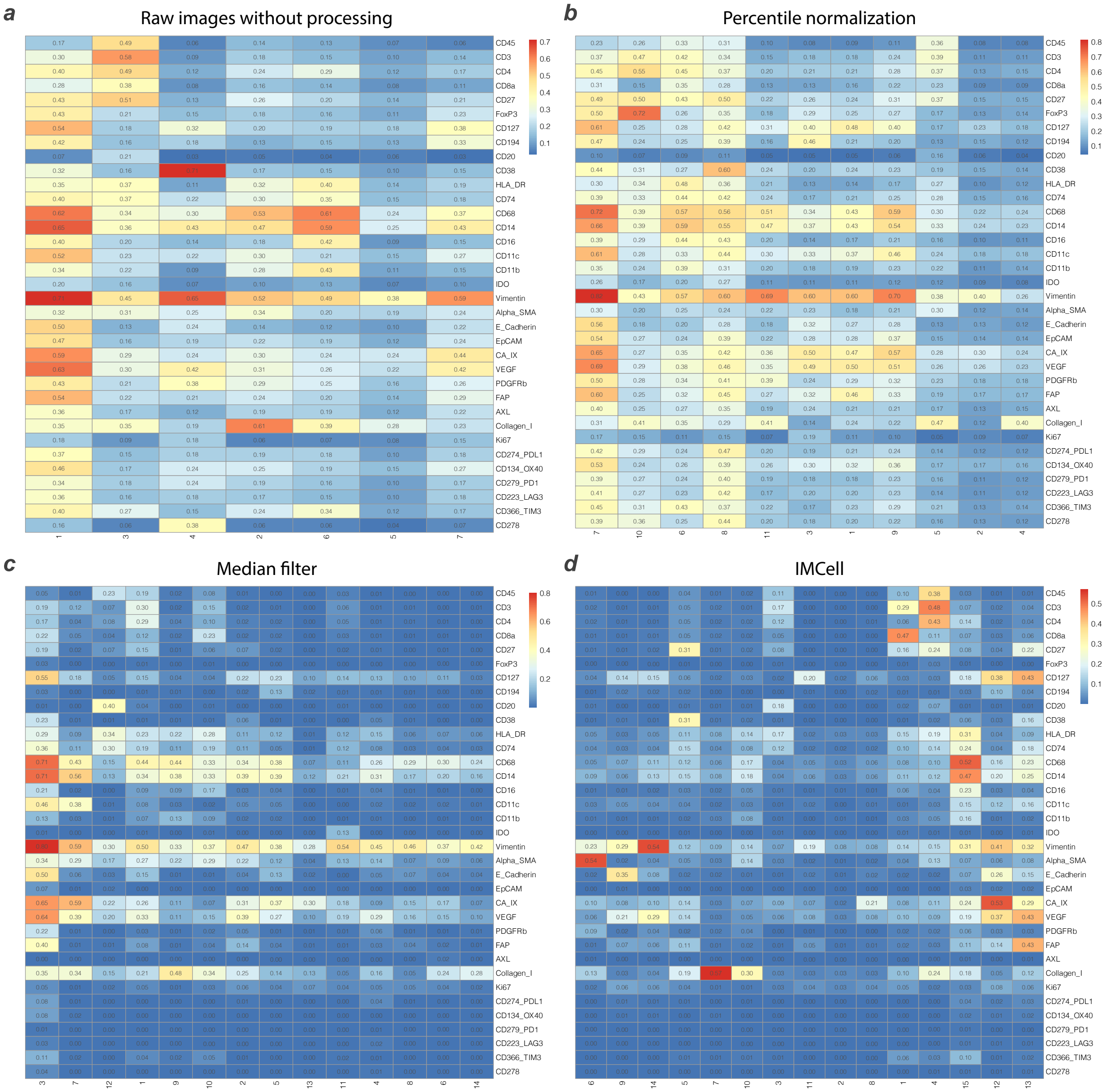
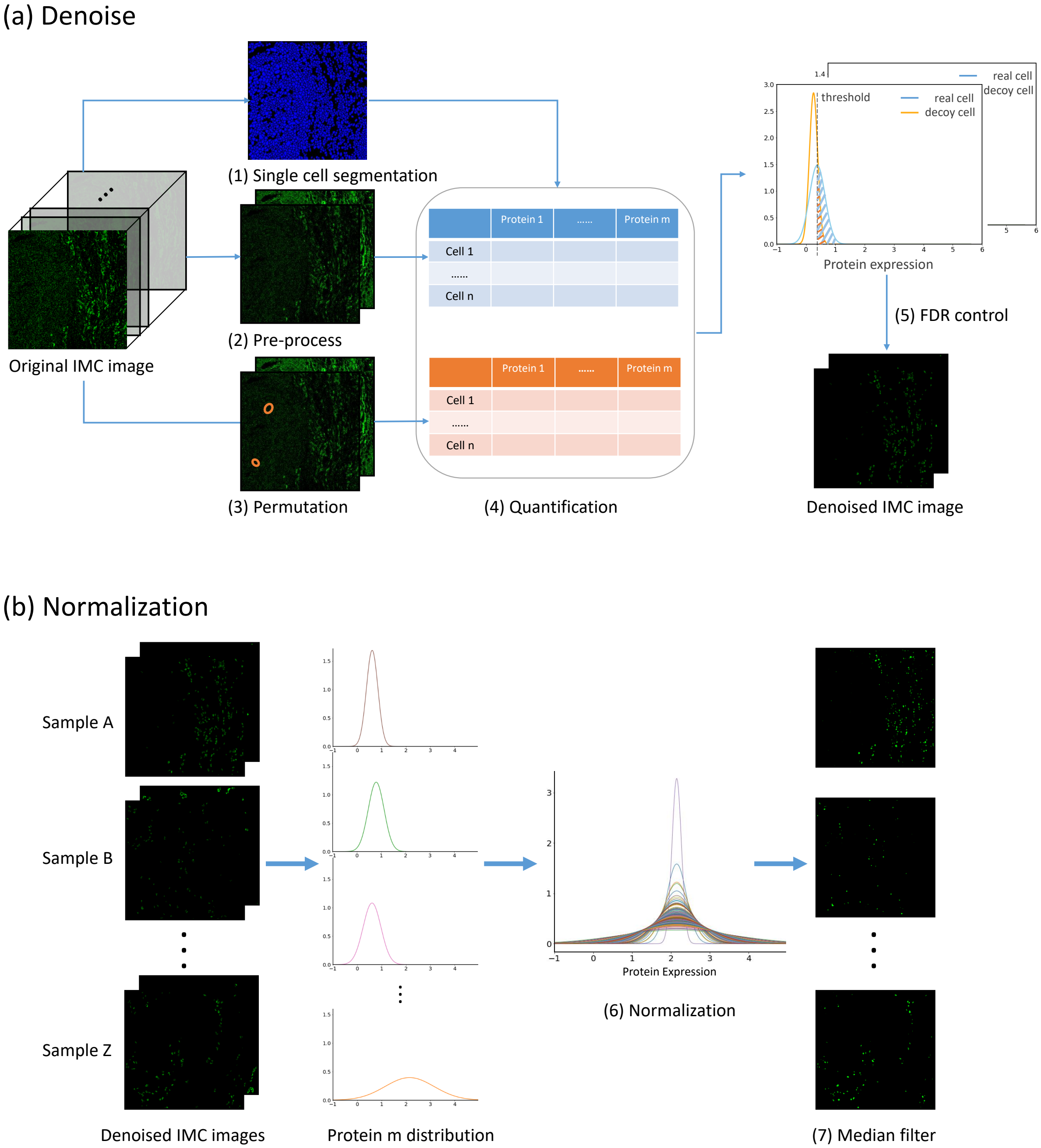


Figure 4: Clustering result from different methods. Heatmap showing mean value of normalized protein expression in each cluster. The high-dimensional single cell expression data were generated from (a) raw IMC images, (b) with 1^{st}-99^{th} percentile method to remove outliers, (c) with median filter to remove salt-and-pepper noise, and (d) with IMCell.



	Protein 1	......	Protein m
Cell 1			
......			
Cell n			

	Protein 1	......	Protein m
Cell 1			
......			
Cell n			

Figure 5: The workflow of IMCell consists of (a) denoising and (b) normalization. The workflow includes the following procedures, (1) single cell segmentation by Dice-XMBD, (2) image pre-processing and hot pixel removal, (3) random generation of decoy cells in potential noise regions, (4) protein quantification for segmented cells and decoy cells, (5) identifying positive cells with FDR control, (6) normalization by scaling using the mean of protein expression of positive cells, and (7) apply the median filter on the denoised and normalized images.