

Semi-Automated Image Analysis Methodology to Investigate Intracellular Heterogeneity in Immunohistochemical Stained Sections

Rifat Hamoudi * Ψ , PhD, CEng, MIEEE
Sharjah Institute for Medical Research
College of Medicine
University of Sharjah
Sharjah, UAE
rhamoudi@sharjah.ac.ae
ORCID ID: 0000-0002-1402-0868

Sarah Hammoudeh Ψ
College of Medicine
University of Sharjah
Sharjah, UAE
sara-hammoudeh@hotmail.com

Arabella Hammoudeh
College of Medicine
University of Sharjah
Sharjah, UAE
arabhammoudeh@hotmail.com

Surendra Rawat
College of Medicine
University of Sharjah
Sharjah, UAE
ssrawat26july@gmail.com

Abstract — The discovery of tissue heterogeneity revolutionized the existing knowledge regarding the cellular, molecular, and pathophysiological mechanisms in biomedicine. Therefore, basic science investigations were redirected to encompass observation at the classical and quantum biology levels. Various approaches have been developed to investigate and capture tissue heterogeneity; however, these approaches are costly and incompatible with all types of samples. In this paper, we propose an approach to quantify heterogeneous cellular populations through combining histology and images processing techniques. In this approach, images of immunohistochemically stained sections are processed through color binning of DAB-stained cells (in brown) and non-stained cells (in blue) to select cellular clusters expressing biomarkers of interest. Subsequently, the images were converted to a binary format through threshold modification (threshold ~ 60%) in the grey scale. The cell count was extrapolated from the binary images using the particle analysis tool in ImageJ. This approach was applied to quantify the level of progesterone receptor expression levels in a breast cancer cell line sample. The results of the proposed approach were found to closely reflect those of manual counting. Through this approach, quantitative measures can be added to qualitative observation of subcellular targets expression.

Keywords—intracellular heterogeneity, image processing, granularity analysis, image thresholding

I. INTRODUCTION

Physiological systems are non-linear and highly complex, performing and regulating numerous functions concomitantly. The complexity of these systems increases exponentially going down the hierarchical levels constituting the human body (e.g. cellular and subcellular levels) [1]. However, until recently, this complexity was not realized by the medical practice; as it focused mainly on targeting higher hierarchical levels (i.e. organs and systems) in diagnosis and treatment. Consequently, an efficient cure for complex chronic diseases (e.g. cancer, asthma, diabetes) could not be achieved.

In the last decade, another magnitude of complexity was introduced revolutionizing the existing knowledge on pathophysiology; the intracellular heterogeneity. Investigations of cellular populations contributing to the structure of various human tissues (e.g. liver, breast, adipose)

revealed highly heterogeneous cellular phenotypes in terms of morphology, genetics, transcriptomics and proteomics [2-5]. These phenotypes develop through various mechanisms including sporadic mutations acquisition, clonal selection, and adaptive transcriptomic reprogramming [6, 7]. The implications of intracellular heterogeneity include disease progression, treatment failure and development of drug resistance [8].

In correspondence to the gravity of these implications, medical practice is currently being redirected to encompass investigations at the tissue, cellular and subcellular levels. Therefore, the demand for approaches to achieve this objective is increasing. Various approaches were developed to capture heterogeneous cellular populations including laser capture microdissection, flow cytometry, microwell platforms (e.g. cyto-seq), microdroplet platforms (Drop-seq and inDrop), microfluidic platforms (Fluidigm C1 single-cell platform), and imaging-based cell sorting platforms (e.g. DEPArray) [9].

Unfortunately the limitations of these techniques hinder their incorporation into the current medical diagnostic practice. One of these limitation is the high cost of the machines and material required to perform such techniques. Moreover, embedding patient sample in paraffin wax is the gold standard for long term preservation; therefore, a significant portion of the accessible samples are received in this form [10]. However, many of the currently existing techniques are incompatible with the formalin-fixed paraffin-embedded (FFPE) samples. Henceforth, there is a high demand on approaches to analyze and capture heterogeneity in these samples. Heterogeneity is evidently observed in immunohistochemically stained sections of patient samples. However, due to the lack of accurate tools to quantify the heterogeneity, it is described using qualitative observations and percentile estimations.

Moreover, intracellular heterogeneity is not limited to patients samples. In vitro models (e.g. primary cells and immortalized cell lines) have been shown to similarly present with intracellular heterogeneity. In vitro investigations are powerful tools in basic science research to investigate various cellular aspects including molecular mechanisms and molecular response to therapy. However, analysis of the in vitro experimental systems conducted traditional molecular

Dr Rifat Hamoudi is funded by Al-Jalila Foundation (Grant code: AJF201741) for this work.

* Corresponding author: Dr. Rifat Hamoudi
email: rhamoudi@sharjah.ac.ae

Ψ Authors contributed equally to this work

methods including western blot and qPCR provides the sum of the average, disregarding the intracellular heterogeneity.

In this paper, we combine histopathology techniques with imaging analysis to investigate intracellular heterogeneity. The approach is based on color classification at the granular level to identify and count cellular structures of interest. This approach can be applied to patient samples as well as aggregated cell line samples.

II. METHODS

A. General methodology outline

The described methodology was applied to images of MCF-7 cell line sections. MCF-7 cells aggregates were created using the thrombin-plasma clotting approach, processed and embedded into paraffin wax in preparation for staining. The sections were stained against estrogen receptor (ER) and progesterone receptor (PR), hormone receptors verified to be over-expressed in the MCF-7 cell line. The expression of the target proteins was visualized with the aid of Diaminobenzidine (DAB)-based immunohistochemistry. Accordingly, cells expressing PR are stained with the brown color of the Diaminobenzidine chromogen; whereas, PR negative present with only the color of the hematoxylin counter-stain (blue).

The images were processed according to the workflow in figure 1. The following image processing and count extrapolation steps were performed using ImageJ (version 1.51).

Analyzing the granularity would allow for inferring the features of the cells from the distribution of the pixels' clusters distinctly from the background noise and staining artifacts.

B. Color Binning

The detection of the stained and non-stained cells for quantifications was conducted according to the statistical color model proposed by Shu et al. [11]. In this model, color detection is based on probabilities of the colors according to the color image histogram. Accordingly, the pixels occupied by each color (stained-brown and non-stained-blue) are quantified and compared against the total number of pixels:

$$Prob(\text{stained pixels}) = \frac{N_{\text{stained pixels within a color bin}}}{N_{\text{stained pixels}}}$$

Where N is the number of pixels.

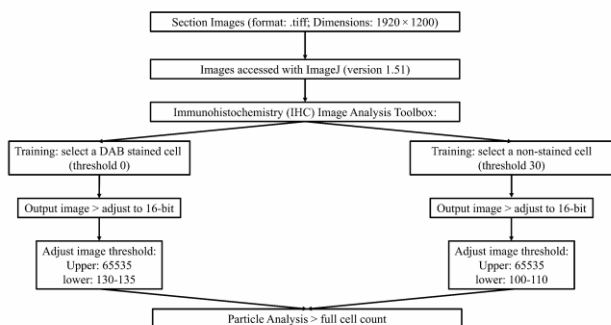


Fig.1. Workflow of the proposed image processing and granularity analysis approach

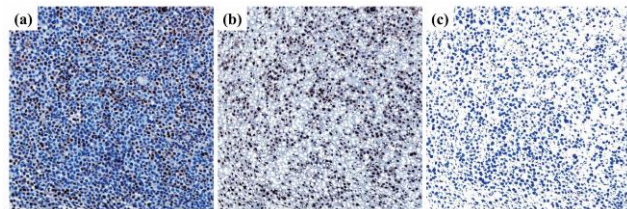


Fig.2. Color Binning. (MCF7 section stained against PR (a) was processed with the color binning tool to capture (b) the cells positively stained against PR in brown and (c) the cells negative for PR expression stained with the haematoxylin counter stain in blue.)

The color data is processed according to the chromaticity rather than luminance. The red-green (rg) and blue-yellow space and the Cb and Cr space from the YCbCr model were adapted to detect the chromaticity signals, with the integration

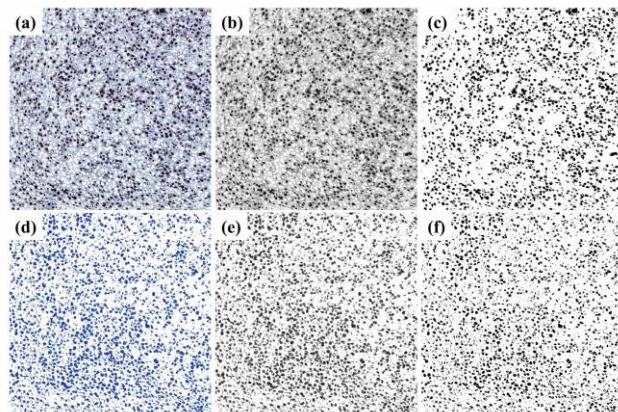


Fig.3. Post-Binning image thresholding. (Images processed with the color binning tool (a,d) are converted to their binary form through changing their type to 16-bit (b,e) and thresholding in the black and white color space (c,f) in preparation for particle analysis.)

of the likelihood ratio for classifying the signals for each of the pixels. In our approach we used the color binning tool in the IHC Image analysis toolbox (ImageJ) to select stained (brown, fig. 2-b) and non-stained cells (blue, fig. 2-c) through the training tool.

C. Thresholding and binary conversion

Subsequently to the color binning (figure 3-a,d), the image type was adjusted to 16-bit in preparation for thresholding (figure 3-b,e). thresholding was adjusted in the black and white (greyscale) color space according to the following specifications: (1) ER/PR positive cells: 130-135 and (2) ER/PR negative cells: 100-110 (figure 3-c,f). The stated threshold values were chosen according to the set of available images; therefore, maximizing the number of visualized cells while minimizing staining artifacts, background, and connections between adjacent cells.

D. Particle Analysis and Cell Count

The cell count was extrapolated from the processed images (color binning and thresholding) using the particle analysis tool in ImageJ. The percentages of the stained and non-stained cells were calculated. Manual cell count was conducted with the use of the cell counter tool in ImageJ. The percentages from both, manual and semi-automatic cell counts were compared to verify the efficiency of the proposed semi-automated counting approach.

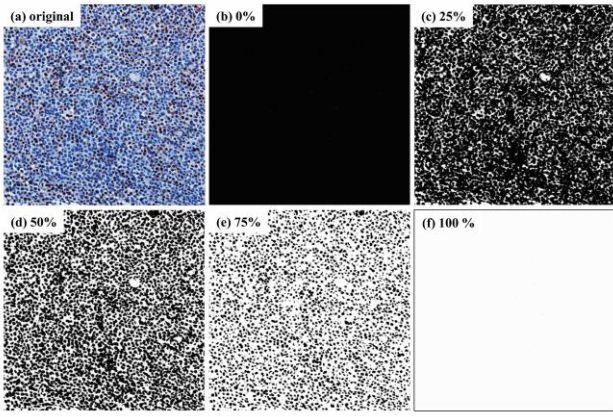


Fig.4. Greyscale thresholding optimization. (*threshold optimized across a range of values including (b) 0, (c) 25, (d) 50, (e) 75, and (f) 100 %.*)

III. EXPERIMENTAL RESULTS

In this paper, we apply the proposed approach to analyze sections of MCF-7 cell line aggregate stained against PR. Initially, we aimed at optimizing the various parameters of the proposed image processing approach (e.g. color binning and thresholding). First, we assessed the efficiency of the thresholding in the greyscale at eliminating the background staining without eliminating cellular structures. The thresholding value was tested across a range incorporating the quartile percentages of the thresholding scale (figure 4). According to the comparisons between the different percentages of thresholding, maintaining the thresholding value within the third quartile provided the best contrast and definition of the cellular structures.

To verify the importance of the color binning for the analysis of the sample granularity, we compared the cell count results from the proposed approach with those of the manual counting and semi-automated counting without binning (only color thresholding in the RGB color space) (figure 5-b,d,f,h). The comparison was conducted on a single image stained against PR.

The cell count percentages extrapolated post color binning reflected those of the manual counting on the contrary to those processed without color binning (figure 5-i). In comparison to the manual cell count, our proposed approach presented with 3.2% false positive rate and 2.3% false negative rate. On the other hand, the semi-automated count of cell post color thresholding only presented with 98.4% false positive rate and 69.6% false negative rate.

Processing the images through modifying color threshold only, resulted in the loss of faintly stained and small cells, therefore, under-estimating the number of PR positive cells. Moreover, it increased the difficulty in distinguishing the background stain from the counter-stain (hematoxylin blue stain), hence, over-estimating the count of the PR negative cells. These shortcomings are overcome through the binning approach as its focused on analyzing color noise and its context in the image.

IV. DISCUSSION

The discovery of Intracellular heterogeneity revolutionized the understating of the human body physiology and pathology. Commonly used molecular methodologies (e.g. western blot, qPCR) yield the sum of the average of

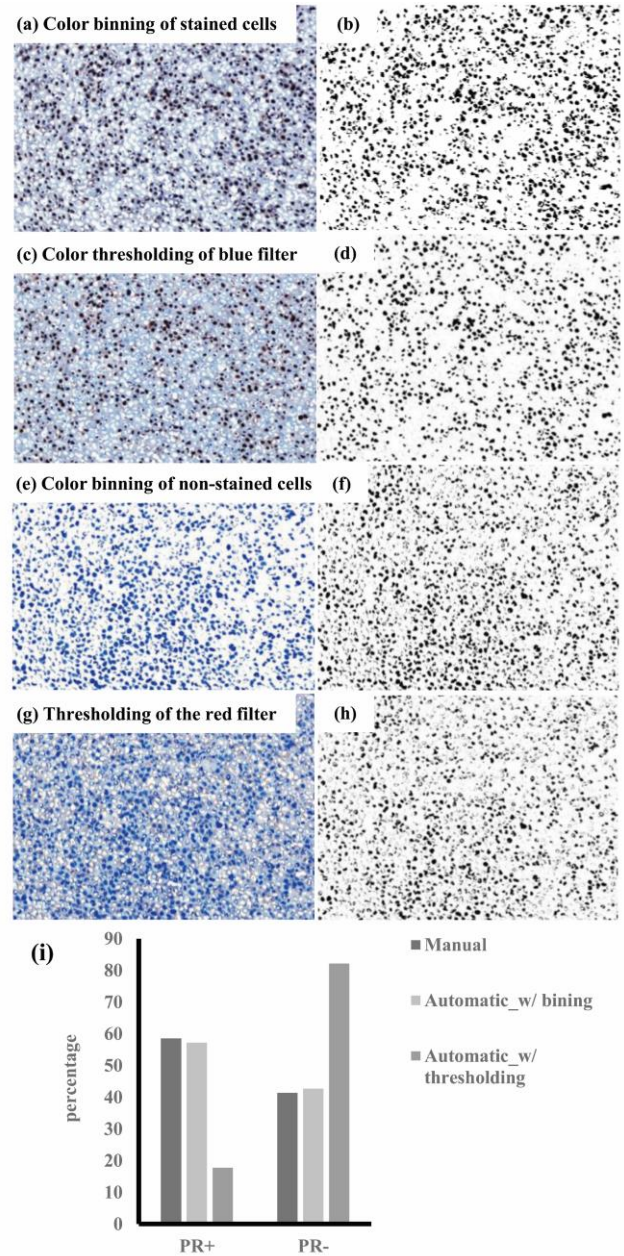


Fig.5. Comparison the proposed approach combining color binning and threshold against thresholding and filter modification only . (*images processed using the color binning method (a,e) are compared against thresholded images in the RGB color space (c,g). All images are then converted to their binary from using the thresholding tool in the black and white color space for particle analysis. (i) quantitative analysis of the processed images using color binning, thresholding and filter modification,*

biomarkers expression under the assumption that these results apply to all the cells in the investigated population. Therefore, the experimental results would lack the input on the heterogeneous subcellular content varying amongst the different cellular clusters within the tissue or population of interest. Henceforth, there is a critical need for approaches to investigate the nature and effects of intracellular heterogeneity.

In this paper, we propose an approach to quantify the heterogeneous expression of molecular targets in immunohistochemically stained sections. As can be seen from the presented results, the proposed approach yields cell count results at a higher accuracy than the traditional thresholding

and filter modification approach. This advantage stems from the color binning step used to identify cellular contents of interest according to color classification and granularity analysis. Verification of the accuracy of the approach was done through comparing the percentages of the cellular populations extrapolated from semi-automated and manual counting. A high similarity was observed between the outputs of both methods. Moreover, these results were consistent across multiple sections stained against two different molecular targets. Therefore, these outputs support the relative accuracy and precision of the proposed approach.

The applications of this approach include the investigation of response to therapy, cellular content of samples of interest, cellular dynamics, and cellular evolution and phenotyping shifting. Above all, this approach can enhance and substantiate the representation of intracellular heterogeneity, providing percentages of the presented cellular phenotypes in addition to qualitative observations.

However, some discrepancies were observed between the semi-automatic and manual counts which could be contributed by some of the approach's limitations. One of these limitations is the limited capacity of the approach to capture all cellular structures due to their irregular shapes and heterogeneous sizes. The proposed approach attempts to linearize a non-linear model (human tissues) resulting in the reduction of cell count accuracy. Moreover, this approach is dependent on the color distribution for the detection of cellular structures. However, the process of the immunohistochemical staining results in color artifacts and nonhomogeneous color scales. Therefore, the analog nature of the color gradient in the sections increases the difficulty of distinguishing the different areas and structures within the sections. Future work will aim to tackle these limitations, develop the methodology to enhance the accuracy of the approach and investigate the developed approach on larger datasets.

V. CONCLUSION

In this paper, we propose an approach combining histopathology techniques with image processing to investigate intracellular heterogeneity. As the proposed approach is based on color binning and analyzing granularity, it does not require the use of training set and neural network analysis to identify cellular features. This approach can aid in semi-automatic but accurate count which will be invaluable for biomedical research including drug discovery. Moreover, this methodology can be adapted to other fields such as remote sensing and geographic feature extraction such as counting different features from satellite images.

ACKNOWLEDGMENT

We would like to thank Al-Jalila Foundation (Grant code: AJF201741) for funding this work.

REFERENCES

- [1] S. Z. Spasić and S. Kesić, "Editorial: Nonlinearity in Living Systems: Theoretical and Practical Perspectives on Metrics of Physiological Signal Complexity," (in English), *Frontiers in Physiology*, Editorial vol. 10, no. 298, 2019-March-29 2019.
- [2] N. Aizarani *et al.*, "A human liver cell atlas reveals heterogeneity and epithelial progenitors," *Nature*, 2019/07/10 2019.
- [3] P. J. Keller *et al.*, "Mapping the cellular and molecular heterogeneity of normal and malignant breast tissues and cultured cell lines," (in eng), *Breast Cancer Res*, vol. 12, no. 5, p. R87, 2010.
- [4] A. E. Cerchiari *et al.*, "A strategy for tissue self-organization that is robust to cellular heterogeneity and plasticity," *Proceedings of the National Academy of Sciences*, vol. 112, no. 7, pp. 2287-2292, 2015.
- [5] Q. Luong, J. Huang, and K. Y. Lee, "Deciphering White Adipose Tissue Heterogeneity," (in eng), *Biology (Basel)*, vol. 8, no. 2, Apr 11 2019.
- [6] D. Stockholm *et al.*, "The origin of phenotypic heterogeneity in a clonal cell population in vitro," (in eng), *PLoS one*, vol. 2, no. 4, pp. e394-e394, 2007.
- [7] R. E. Hynds, E. Vladimirov, and S. M. Janes, "The secret lives of cancer cell lines," *Disease Models & Mechanisms*, vol. 11, no. 11, p. dmm037366, 2018.
- [8] M. Greaves, "Evolutionary Determinants of Cancer," *Cancer Discovery*, vol. 5, no. 8, pp. 806-820, 2015.
- [9] F. Ye, W. Huang, and G. Guo, "Studying hematopoiesis using single-cell technologies," (in eng), *Journal of hematology & oncology*, vol. 10, no. 1, pp. 27-27, 2017.
- [10] S. Magdeldin and T. Yamamoto, "Toward deciphering proteomes of formalin-fixed paraffin-embedded (FFPE) tissues," (in eng), *Proteomics*, vol. 12, no. 7, pp. 1045-1058, 2012.
- [11] J. Shu, G. Qiu, M. Ilyas, and P. Kaye, "Biomarker Detection in Whole Slide Imaging based on Statistical Color Models," *The MIDAS Journal-Computational Imaging Biomarkers for Tumors (CIBT)*, 2010.