Automated Segmentation and Measurement for Cancer Classification of HER2/neu Status in Breast Carcinomas

Lee Sing CHEONG[1] and Angela JEAN[2]
Bioinformatics Group
Nanyang Polytechnic
Singapore
Bioinformatics Research Centre
Nanyang Technological University
Singapore
[1]e-mail: Alice_CHEONG@nyp.gov.sg
[2]e-mail: AngelaJEAN@gmail.com

Tsu Soo TAN[3] and Waiming KONG[4]
Bioinformatics Group
Nanyang Polytechnic
Singapore
[3]e-mail: TAN_Tsu_Soo@nyp.gov.sg
[4]e-mail: KONG_Wai_Ming@nyp.gov.sg

Soo Yong TAN
Department of Pathology
Singapore General Hospital
Singapore
e-mail: tan.soo.yong@sgh.com.sg

***Abstract*—The HER2/neu protein is over-expressed in 20%-30% of breast cancer cases, and is significantly associated with increased breast cancer recurrence and worse prognosis. The assessment of HER2 protein level is visualized using immunohistochemistry (IHC) assays, which is subjected to inter-observer and intra-observer variability. In this paper, we reduce variability by an automated segmentation and measurement system for IHC-stained breast tissue images. From the dataset of breast tissue images, the system is able to obtain the nuclei and the orange stained cell membranes of the cells, quantify the continuous orange hue of the cell membranes, and identify nuclei that are bounded by orange stained cell membrane. This system also suggests a putative assessment classification score for each image based on the same assessment protocol specified for histopathologist. Using the dataset of 42 clinically IHC-scored images, the system correctly suggested the corresponding putative assessment classification score for 39 of the images, achieving an accuracy of 92%.**

Keywords - HER2; neu; breast carcinoma; automated segmentation and measurement; cancer classification.

I. INTRODUCTION

HER2 is the abbreviation for "Human Epidermal growth factor Receptor 2", while oncogene neu was named after the cell line being derived from a rodent glioblastoma cell line, which is a type of neural tumor. It is also named as ErbB-2 for its similarity to ErbB (avian erythroblastosis oncogene B).

The HER2/neu is a cell membrane surface-bound receptor tyrosine kinase protein and is a member of the human epidermal growth factor receptor (HER) family. It interacts with other HER receptors to regulate cell growth, differentiation and survival [1].

HER2/neu protein is over-expressed in 20%-30% of breast cancer cases [2,3], and is significantly associated with increased breast cancer recurrence and worse prognosis [4]. It is the target of the monoclonal antibody trastuzumab (Herceptin US brand name) [5,6], and is FDA approved as part of a treatment plan for the adjuvant treatment of HER2 positive breast cancer patient [6].

In this paper, we will cover the current practice in HER2 assessment, the dataset used, the method employed, the results obtained and discuss the significant observations.

II. HER2 ASSESSMENT

HER2 protein's level of expression is measured through the use of immunohistochemistry (IHC) assays. In an IHC assay, the antibodies bind to the antigen (HER2/neu protein) in the biological tissue.

The assessment protocol assigns a score of 0 to a sample tissue when no observable staining occurs or when less than 10% of the membrane of tissue cells is stained. A score of 1 is assigned when greater than 10% of the membrane of tissue cells is faintly or barely perceptibly stained. A score of 2 is assigned when greater than 10% of the complete membrane of tissue cells is weakly or moderately stained. Lastly, the maximum score of 4 is assigned when greater than 10% of the complete membrane of tissue cells is strongly stained.

While a protocol is in place to assess biological tissue according to membrane staining, inter-observer and intra-observer variability exist. The inter-observer variability reflects the systematic differences among the observers, while the inter-observer variability reflects the discrepancy

 ISBN: 978-1-61208-137-3

resulting from the use of human perception to classify the continuous hue of staining and the completeness of enclosure of the membrane staining.

Thus, an automated segmentation and measurement method that quantify the classification of continuous hue of staining and the completeness of enclosure of the membrane staining, will relieve the histopathologist from performing quantitative classification, and instead concentrates on the assessment of the tissue sample using expert knowledge. This will reduce the amount of assessment variability, and decrease the time used to assess each tissue sample.

III. DATASET

The dataset used in this paper consist of 42 IHC-stained and clinically IHC-scored breast tissue histological images, which were obtained from Dr Tan Soo Yong. Of these 42 images, 3 images were of IHC score 0, 12 images were of IHC score 1, 11 images were of IHC score 2 and 16 images were of IHC score 3.

Each of these images is of dimensions 1300 pixels by 1030 pixels and is in the TIFF image format. They are manually acquired through microscopic imaging of normal breast tissue and varying grades of breast cancer tissue that had been IHC-stained and mounted on histological slides.

For the IHC-stained normal breast tissue cells, the nuclei are stained blue-purple and the cell membranes are hardly stained. For IHC-stained abnormal breast tissue cells, the nuclei are also stained blue-purple, while the cell membranes are stained in varying orange hue.

IV. METHOD

Here, we present an automated segmentation and measurement system that is able to identify nuclei that are bounded by orange stained cell membrane and quantify the continuous hue of orange stained cell membrane. In addition, based on the assessment protocol specified for histopathologist, but without any additional expert knowledge that is accumulated through experience for an expert, a putative assessment classification score is suggested.

A. *Identify orange hue in tissue image*

The breast tissue images are identified as either containing cell membranes that are not stained orange or are stained orange in the following steps:

Step 1: The color space of the image is converted from RGB color space to CIE-Lab color space.

Step 2: The frequency of pixels having each of the possible chromaticity (chromatic component a and chromatic component b) is calculated.

Step 3: The extremal chromaticity maxima values of the frequency of pixels are obtained.

Step 4: The image is classified as breast tissue containing cell membranes that are without orange stain if there are two extremal chromaticity maxima points (It can be seen from Figure 1 that there is two extremal chromaticity maxima points in a plot of maxima chromaticity points for breast tissue image without orange stained cell membranes).

Likewise, it is classified as breast tissue containing cell membranes that are with orange stain if there are three extremal chromaticity maxima points (It can be seen from Figure 2 that there is three extremal chromaticity maxima points in a plot of maxima chromaticity points for breast tissue image with orange stained cell membranes).

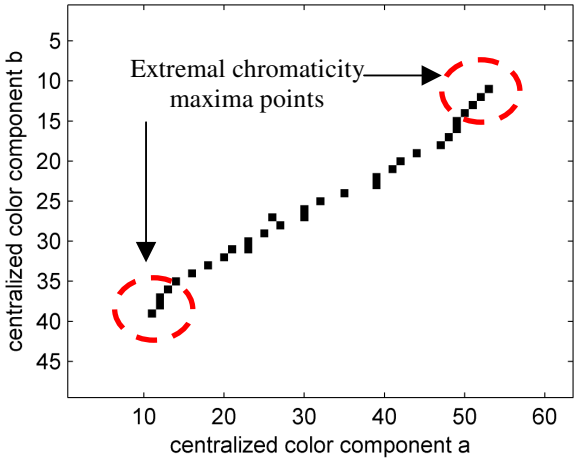

Figure 1. Plot of maxima chromaticity points for breast tissue image without orange stained cell membranes

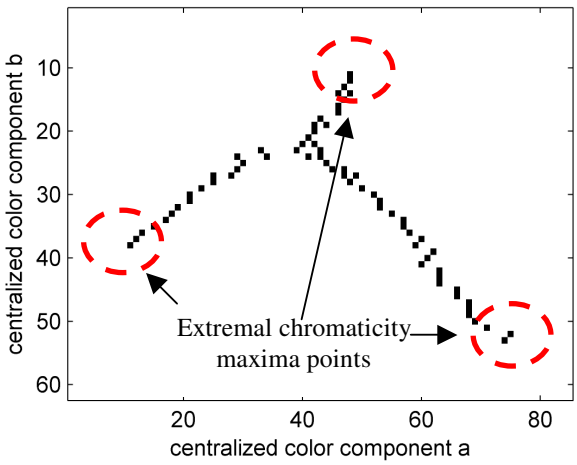

Figure 2. Plot of maxima chromaticity points for breast tissue image with orange stained cell membranes

B. *Segment breast tissue image with orange hue*

The images identified as breast tissue image with orange stained cell membranes is further processed in the following sections. In this section, the breast tissue image with orange stained cell membranes is segmented to obtain the nuclei cluster and the orange stained cell membranes cluster in the following steps:

 ISBN: 978-1-61208-137-3

Step 1: The chromaticity value corresponding to the maximum frequency of chromaticity of the pixels is obtained.

Step 2: The chromaticity value corresponding to the average chromaticity of the maximum frequency's chromaticity obtained in the previous step and the extremal chromaticity maxima value corresponding to the orange hue is obtained.

Step 3: The three extremal chromaticity maxima values obtained in the previous section, the maximum frequency's chromaticity value obtained in Step 1, and the average chromaticity value obtained in Step 2, are used as seed values to perform k-means clustering on the image.

Step 4: The cluster corresponding to the k-means centroid with least color-component b is classified as the nuclei cluster, and is shown in Figure 3.

Step 5: The two clusters corresponding to the k-means centroids with the greatest two values of the sum of the color-components are both classified as the orange stained cell membranes cluster, and is shown in Figure 4.

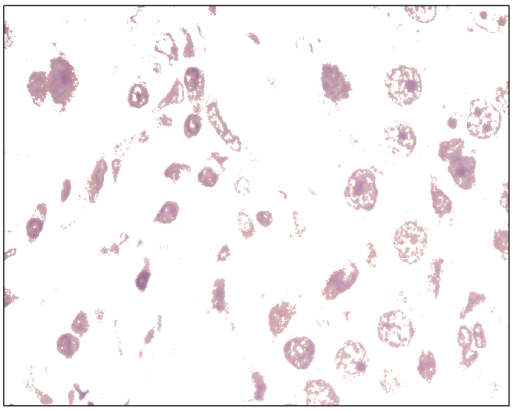

Figure 3. The cluster classified as nuclei cluster

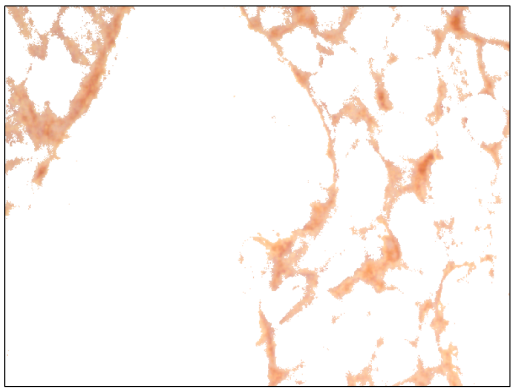

Figure 4. The cluster classified as orange stained cell membranes

C. *Refine the nuclei cluster*

The nuclei cluster, being identified by k-means clustering, contains stray pixels and pixel regions that do not belong to any nucleus. Based on each of the connected region's measurement, the nuclei cluster is further refined by the following steps:

Step 1: The nuclei cluster mask of the nuclei cluster is obtained.

Step 2: Morphological opening is performed on the nuclei cluster mask to remove stray pixels.

Step 3: Each connected region of the nuclei cluster mask, based on neighboring 8-connectivity measurement is individually labeled.

Step 4: For each of the connected region, based on the regional measurement properties of Euler number, major axis, minor axis length and filled area, each connected region is classified as being a nucleus region or not a nucleus region.

D. *Identify orange stained cell membrane bounded nucleus*

The region surrounding each of the connected nucleus region that had been obtained in the previous section of nuclei cluster refinement, is inspected to identify whether it is bounded by orange stained cell membrane in the following steps:

Step 1: For each connected nucleus region obtained in the previous section, its connected nucleus region mask is obtained.

Step 2: Morphological dilation of the connected nucleus region mask is performed to obtain the extended region mask that includes both the connected nucleus region and the surrounding region.

Step 3: The cell membrane region mask is obtained by identifying pixels that are in the orange stained cell membranes cluster obtained by k-means in the earlier section and is within the extended region mask obtained in the previous step.

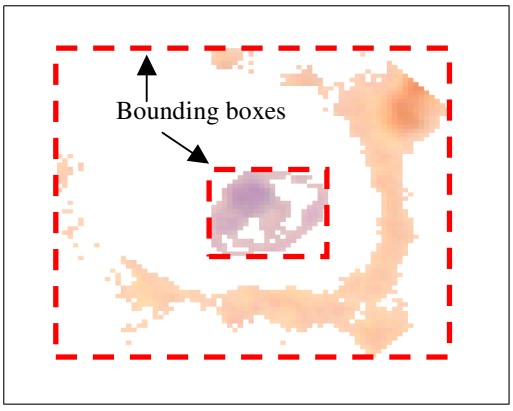

Figure 5. The bounding boxes of the nucleus region mask and the cell membrane region mask for a nucleus that is bounded by orange stained cell membrane

 ISBN: 978-1-61208-137-3

Step 4: The regional measurement property of bounding box for both the nucleus region mask and the cell membrane region mask are obtained.

Step 5: The nucleus is classified as bounded by the orange stained cell membrane if the bounding box of the cell membrane region mask is enclosing the bounding box of the nucleus region mask, and is shown in Figure 5.

E. Quantify continuous hue of orange stain cell membrane

In the following steps, the continuous hue of orange stain for each orange stained cell membrane bounded nucleus is quantify into three ranges; The first range quantifies the description of faintly or barely perceptibly stained cell membrane, the second range quantifies the description of weakly or moderately stained cell membrane, and the third range quantifies the description of strongly stained cell membrane.

Step 1: Using the cell membrane region mask obtained in the previous section, the mean CIE lightness of the orange stained cell membrane region is obtained.

Step 2: Using thresholding, the mean CIE lightness for each orange stained cell membrane bounded nucleus is quantified into one of the possible three ranges corresponding to (1) faintly or barely perceptibly stained membrane, (2) weakly or moderately stained membrane or (3) strongly stained membrane.

F. Provide putative assessment classification score

Based on the percentage of orange stained cell membrane bounded nuclei for each range over the total number of nuclei, a putative assessment classification score is provided for each of the image by the following steps:

Step 1: The number of orange stained cell membrane bounded nuclei for each range is obtained using the quantification of each orange stained cell membrane bounded nucleus identified in the previous section.

Step 2: The total number of connected nucleus regions is obtained based on neighboring 8-connectivity measurement of the refined nuclei cluster obtained in the earlier section.

Step 3: The percentage of the number of orange stained cell membrane bounded nuclei for each range obtained in Step 1 over the total number of nuclei obtained in Step 2 is obtained.

Step 4: The tissue image containing cell membranes that are not stained orange or consists of less than 10% of the orange stained cell membrane bounded nuclei in the range corresponding to faintly or barely perceptibly stained membrane is given a putative assessment classification score of 0.

The tissue image that consists of greater than 10% of the orange stained cell membrane bounded nuclei in the range corresponding to faintly or barely perceptibly stained membrane is given a putative assessment classification score of 1.

The tissue image that consists of greater than 10% of the orange stained cell membrane bounded nuclei in the range corresponding to weakly or moderately stained membrane is given a putative assessment classification score of 2.

The tissue image that consists of greater than 10% of the orange stained cell membrane bounded nuclei in the range corresponding to strongly stained membrane is given a putative assessment classification score of 3.

G. Visualize the orange hue quantification

Visualization of the identified nuclei and the quantification of the continuous hue of orange of the cell membrane bounded nuclei are provided by the use of four different colors. In the following steps, four different colors are used to correspondingly outline nuclei's whose cell membranes are (1) not stained, (2) faintly or barely perceptibly stained, (3) weakly or moderately stained, or (4) strongly stained.

Step 1: The perimeter of each of the nucleus identified in the refined nuclei cluster is obtained.

Step 2: Based on the quantification of each of the orange stain cell membrane bounded nucleus, the perimeter of these nuclei are colored with one of the three colors according to their quantification of orange hue.

Step 3: The remaining nuclei whose perimeter are colored in the Step 2, are those whose cell membranes are not stained. They are colored a different color from the three colors used in Step 2.

V. RESULT

The putative assessment classification score suggested by the automated segmentation and measurement system described in this paper achieved an accuracy of 92% with 39 out of 42 images in the dataset having the same corresponding classification score as the IHC score provided by clinician.

The corresponding putative assessment classification score for 3 out of 3 images of IHC score 0, 10 out of 12 images of IHC score 1, 10 out of 11 images of IHC score 2 and 16 out of 16 images of IHC score were correctly suggested.

For the 2 images of IHC score 1 with incorrectly suggested classification score, they were both scored as one category more serious with classification score 2 that corresponds to IHC score 2. Likewise, for the single image of IHC score 2 with incorrectly suggested classification score, it is scored as one category more serious with classification score 2 that corresponds to IHC score 3.

Thus, the automated segmentation and measurement system provides visualization of the identified nuclei and the quantification of the continuous hue of orange of the cell membrane bounded nuclei, along with a suggestion of putative assessment classification score that is highly accurate.

VI. DISCUSSION

In this section, five observations are being discussed.

 ISBN: 978-1-61208-137-3

A. *Challenge of variable stains' hue*

There is great variance in the stains' hue for the same components of cells in the tissue images. As seen in Figure 6, the same description of blue-purple stain has different chromaticity in different image.

These variances in hue can be due to many reasons, such as the difference in the amount of stains used or amount of washing performed during the process of preparing of the IHC-stained histological slides, the different lighting condition of the microscopy and the difference in automatic white balancing of the image-capturing device.

The stain for the same cell component in each image having different chromaticity pose the difficulty of requiring the identification of the number of hues in the image, and the identification of the nuclei cluster and the cell membranes cluster, to be non-dependent on exact chromaticity and has to be adaptive to each of the image.

In this paper, the identification of the number of hues in the image circumvent this difficulty through the use of the chromaticity frequency matrix of the individual image to obtain the number of extremal chromaticity maxima points, which corresponds to the same number of hues in the image.

Likewise, k-means clustering method used to identify the nuclei cluster and the cell membranes cluster is able to circumvent this difficulty, as the method is not dependent exact chromaticity.

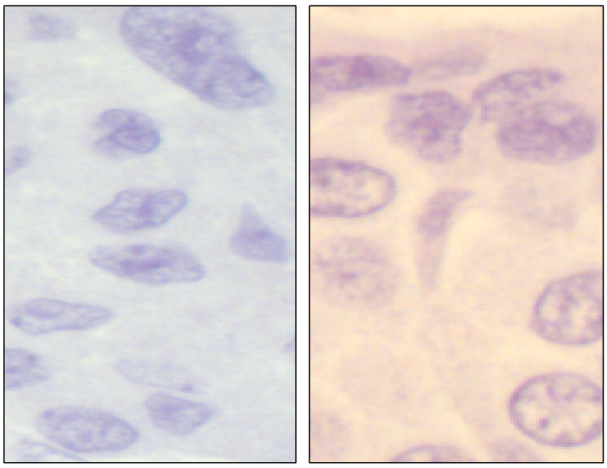

Figure 6. Variance in stains in different image

B. *Challenges in using k-means clustering*

The challenges in using k-means cluster are that the number of seed values has to be known prior and the provision of initial seed values greatly influence the clustering result.

The first challenge of knowing the number of seed values in advance had been solved in the previous section through the use of the chromaticity frequency matrix of the individual image to obtain the number of extremal chromaticity maxima points, which corresponds to the same number of hues in the image.

The second challenge of providing seed values that converge to the required solution is solved through the use of seed values adapted to the individual image, which includes the three extremal chromaticity maxima points and two additional intermediate maxima points.

C. *Reliance on cell characteristics domain knowledge in the refinement of nuclei cluster*

The refinement of nuclei cluster relied on the cell characteristics domain knowledge, in which each nucleus is required to have a minimum size and has the spherical-like shape. This cell characteristics information varies depending on the type of cells that are contained within the breast tissue image.

D. *Challenge of identifying incompletely stained cell membrane as enclosing nucleus*

The incompletely stained cell membrane of nucleus in any of the image can consist of multiple disconnected stained regions of varying distance from the closest point of the nucleus. This poses a challenge in identifying whether each nucleus is being bounded by orange stained cell membrane.

In this paper, in order to overcome the challenge, the bounding box of possibly disconnected regions of the cell membrane pixels that are within an extended distance from the nucleus is evaluated on the criteria of whether it encloses the bounding box of the nucleus under inspection.

The use of bounding box relax the criteria from identifying the entire surrounding cell membrane, which might be incompletely stained, to identifying sufficient number of pixels of the cell membrane to form the bounding box that will enable correct solution to be obtained.

E. *Reliance on domain knowledge of cell membranes staining in quantifying cell membranes' stain*

The quantifying of cell membranes' stain into one of the possible three ranges corresponding to (1) faintly or barely perceptibly stained membrane, (2) weakly or moderately stained membrane or (3) strongly stained membrane rely on domain knowledge of the threshold values for the mean CIE lightness.

F. *Reliance on the assessment protocol specified for histopathologist to suggest putative assessment classification score*

The suggested putative assessment classification score rely on the assessment protocol by the scoring of each image based on the percentage of orange stained cell membrane bounded nuclei for each of the three ranges over the total number of nuclei.

VII. CONCLUSION AND FUTURE WORK

We have presented an automated segmentation and measurement system that is able to identify nuclei that are bounded by orange stained cell membranes, quantify the

 ISBN: 978-1-61208-137-3

continuous hue of orange stained cell membrane, and suggest a putative assessment classification score for each image.

Using the dataset of 42 clinically IHC-scored images, the system correctly suggested the corresponding putative assessment classification score for 39 of the images, achieving an accuracy of 92%.

Of the 2 IHC-scored 1 images and the single IHC-scored 2 image, whose score are incorrectly suggested, the putative assessment classification score suggested were one category more serious.

In addition, we discussed the challenges posed by the nature of the images, the challenges of the method being used, and the reliance of the method used on domain knowledge and assessment protocol.

As the system was tested on 42 clinical IHC-scored images, we propose that future work includes testing on a larger dataset.

ACKNOWLEDGEMENT

This study was funded by the Ministry of Education, Singapore (Innovation Fund grant no. MOE2008-IF-1-018).

REFERENCES

[1] Yarden Y., Biology of HER2 and Its Importance in Breast Cancer, Oncology, Vol. 61, Suppl. 2, Pg. 1-13, 2001.

[2] McCann A. H., Dervan P. A., O'Regan M., Codd M. B., Gullick W. J., Tobin B. M., and Carney D. N., Prognostic Significance of c-erbB-2 and Estrogen Receptor Status in Human Breast Cancer, Cancer Research, Vol. 51, Pg. 3296-3303, 15 June 1991.

[3] Slamon D. J., Godolphin W., Jones L. A., Holt J. A., Wong S. G., Keith D. E., Levin W. J., Stuart S. G., Udove J., Ullrich A., and al. et., Studies of the HER-2/neu Proto-oncogene in Human Breast and Ovarian Cancer, Science, Vol. 244, Issue 4905, Pg. 707-712, 12 May 1989.

[4] Salmon D. J., Clark G. M., Wong S. G., Levin W. J., Ullrich A., and McGuire W. L., Human Breast Cancer: Correlation of Relapse and Survival with Amplification of the HER-2/neu Oncogene, Science, Vol. 235, No. 4785, Pg. 177-182, 9 January 1987.

[5] Albanel J., Codony J., Rovira A.., Mellado B., and Gascon P., Mechanism of Action of Anti-HER2 Monoclonal Antibodies, Advances in Experimental Medicine and Biology, Vol. 532, Pg. 253-268, 2003.

[6] Goldenberg M. M., Trastuzumab, A Recombinant DNA-derived Humanized Monoclonal Antibody, A Novel Agent for the Treatment of Metastatic Breast Cancer, Clinical Therapeutics, Vol 21, Issue. 2, Pg.309-318, February 1999.

 ISBN: 978-1-61208-137-3