Image Processing Pipeline to Compute Homologous Recombination Score

Haswanth Vundavilli*
Department of Electrical and Computer Engineering & The Center for Bioinformatics and Genomic Systems Engineering, Texas A&M University, College Station, USA

Manuela Tumiati
Research Program in Systems Oncology, Faculty of Medicine, University of Helsinki, Finland

Sampsa Hautaniemi
Research Program in Systems Oncology, Faculty of Medicine, University of Helsinki, Finland

Aniruddha Datta
Department of Electrical and Computer Engineering & The Center for Bioinformatics and Genomic Systems Engineering, Texas A&M University, College Station, USA

Liisa Kauppi
Research Program in Systems Oncology, Faculty of Medicine, University of Helsinki, Finland

ABSTRACT

DNA double-strand breaks (DSBs) occur frequently in eukaryotic cells, and the homologous recombination pathway (HR) is one of the major pathways required to repair these breaks. However, tumor cells that are able to repair DSBs are unlikely to die due to damage incurred by DNA damaging chemotherapies, such as platinum compounds. While platinum-based therapies have been effective in treating various cancers, they also carry harsh side effects, and thus ideally platinum should be used when the probability of treatment resistance is low. HR scores provide a measure for patients' tumor's HR capacity and have been shown to predict their chemotherapy response and long-term survival. Calculating this score manually from immunofluorescence microscopy images for each patient is error-prone and time-consuming. Herein, we propose an image processing pipeline that takes as input imaging data from three emission channels (representing nuclei, S-phase cells, and HR-mediated repair in a tumor slice) from an epifluorescence microscope and computes the HR score. Our open-source methodology forms a rationale to develop similar approaches in predicting chemotherapeutic responses and facilitating to make treatment decisions.

CCS CONCEPTS

• **Computing methodologies** → Computer graphics; Image manipulation; Image processing.

*Email: hashwanthvv@tamu.edu

ICBET 2022, April 20–23, 2022, Tokyo, Japan

ACM ISBN 978-1-4503-9577-9/22/04…$15.00
https://doi.org/10.1145/3535694.3535704

KEYWORDS

Image Processing, DNA double-strand breaks, Homologous recombination, Platinum chemotherapy

ACM Reference Format:
Haswanth Vundavilli, Manuela Tumiati, Sampsa Hautaniemi, Aniruddha Datta, and Liisa Kauppi. 2022. Image Processing Pipeline to Compute Homologous Recombination Score. In *2022 12th International Conference on Biomedical Engineering and Technology (ICBET) (ICBET 2022), April 20–23, 2022, Tokyo, Japan.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3535694.3535704

1 INTRODUCTION

In platinum-based chemotherapy, platinum compounds, such as cisplatin, oxaliplatin, or carboplatin, are used alone or in combination with other drugs to kill cancer cells [1]. These drugs have been notable for treating different human cancers including lung, ovarian, and bladder cancer. In high-grade serous ovarian cancer (HGSC) current first-line therapy consists of surgery and platinum-based chemotherapy. While platinum has increased overall survival of the patients, there are some patients who do not respond to platinum or response is only for few months. One of the dominant platinum resistance mechanisms in HGSC is DNA repair mechanisms and it would be important to identify patients who are in high risk of poor response to platinum due to DNA repair mechanisms.

Cells must constantly deal with DNA lesions such as double-strand breaks (DSBs) from both internal and external sources. These lesions, if not repaired, can result in cell death or if the cell survives, result in genomic instability [2]. Homologous recombination (HR) pathway is a key mechanism for repairing DSBs [3]. Platinum-based drugs operate by forming highly reactive platinum complexes that ultimately generate DSBs [4]. Elimination of platinum-induced DSBs requires HR repair. Hence, analyzing the activity of the HR pathway is one of the means to predict a patient's response to platinum chemotherapy.

Homologous recombination deficiency (HRD) scores quantify the activity of HR and have been shown to successfully predict chemotherapeutic responses in multiple cancers. Using a HR deficiency (HRD) score based on genomic features ("genomic scars"), Telli et al. (2016) identified HR-deficient tumors and responses to platinum chemotherapy in patients with triple-negative breast cancer [5]. Tumiati et al. (2018) proposed a functional HR score, based on measuring HR capacity ex vivo, that was able to predict primary platinum sensitivity and long-term survival in ovarian cancer patients with high statistical significance [6]. Tumor HRD is a key factor for patient's survival and response to platinum chemotherapy in TNBC and ovarian cancer.

The recent advent of data mining techniques and machine learning methodologies have revolutionized biomedical research by reducing subjective bias and human-caused errors [7]. Specifically, image processing has aided in analyzing microscopy images of cells and tissues with high accuracy [8]. In cancer research, often hundreds of microscopy images are generated from tumor specimens from patients in a cohort. Analyzing these images manually would be both tedious and subject to human bias. Consequently, a number of image processing tools and algorithms have been developed to segment and analyze for example histopathology slides [9], [10], [11]. A manual approach to compute HR scores from microscopy images of tumor samples is clearly inefficient and prone to errors. Herein, we suggest an image processing pipeline to compute HR scores automatically and efficiently.

Image processing techniques generally involve five basic components: a) image acquisition, b) image pre-processing, c) image segmentation, d) image post-processing and e) image analysis. Image segmentation is the most critical step for any technique to perform well. There are a large number of effective segmentation methods, but there is no single tool that is effective on diverse applications. However, watershed algorithms have been shown to work effectively in segmenting cells in microscopy images that are densely crowded [12]. Hence, using a modified watershed algorithm and using object properties of different cells, we developed an image processing pipeline to compute the HR score proposed by Tumiati et al. [6].

2 METHODOLOGY

We now discuss our step-by-step methodology using the five components of image processing and then evaluate its performance. Images were two-dimensional captures of 4-micrometer sections from formalin-fixed, paraffin-embedded tumor specimens from HGSC patients [6]. Slides with tissue sections underwent immuno-incubation with primary antibodies of interest, followed by incubation with immunofluorescently labeled secondary antibodies. DAPI (4', 6-diamidino-2-phenylindole), a fluorescent stain that binds to DNA, was applied to the slides to detect nuclei. Slides were scanned at 20x on a Pannoramic 250 FLASH II slide scanner (3DHISTECH), and high resolution .tiff images were generated for the three separate channels.

2.1 Primary and secondary antibodies used

HR repair can only be employed in the presence of the sister chromatid, that is during S phase of the cell cycle. Therefore, a marker for S phase cells is required to identify those tumor cells in the correct time window for HR. Geminin is a protein that is selectively expressed during the S-M phase of the cell cycle. A third marker required is RAD51, a protein involved in a key step of HR. Presence of RAD51 indicates HR proficiency, whereas absence of RAD51 indicates HR deficiency. Tumor specimens were exposed to DNA damage ex vivo using ionizing radiation, fixed and sectioned. Tumors sections are stained with DAPI, and with antibodies against geminin and RAD51. Each of these markers is detected in a separate emission channel.

2.2 Image pre-processing

After image acquisition, we pre-process the images to remove noise and binarize them. Dust or contaminating oil particles on the objective front lens often lead to bright spots or aberrations in microscopic images. We used a basic pre-processing technique and filtered out portions that were too bright and removed any unusual intensities.

2.3 Image segmentation

Image segmentation means partitioning a digital image into multiple regions. There have been multiple image segmentation techniques developed in this vein and the watershed algorithm has proved to be a powerful tool in applications such as cell segmentation where the objects (cells) are close to each other. The underlying principle of the watershed algorithm comes from geography. Suppose we have a topographic surface with hills (local maxima) and valleys (local minima). We fill up the surface with water starting from local minima, and whenever we encounter a local maximum, we build a barrier (segmentation). We repeat this process until the water has reached the highest peak and the resultant barriers correspond to our segmented image.

The traditional watershed algorithm is not completely reliable with cell segmentation as it is prone to over-segment cells whenever it encounters too many local maxima. Hence, we used a modified watershed algorithm and filtered out small local maxima and filled the holes inside the cells. This technique corrected the segmentation error and significantly improved the performance.

2.4 Image post-processing to distinguish tumor cells from stromal and immune cells

To compute the HR score, we must first identify the tumor cells within the tissue section, which contains also other cell types, mainly stromal cells and immune cells. To do so, we exploited the fact that DAPI-stained appearance of each of these cell types has distinct features (object properties).

After image segmentation, we computed various object properties such as area, eccentricity, and pixel density for each cell (tumor, stromal cells, immune cells). Then, using a supervised learning approach, we learned the thresholds of different object properties and classified tumor, stromal cells, and immune cells separately. For example, immune cells glow brightly and can be filtered out with an intensity filter. Similarly, the stromal cells are rectangular and can be filtered out using an eccentricity filter.

2.5 Calculating HR score

We quantify HR capacity (= HR score) by estimating the fraction of tumor cells that are successfully repairing their DNA damage to the total number of tumor cells undergoing DNA replication.

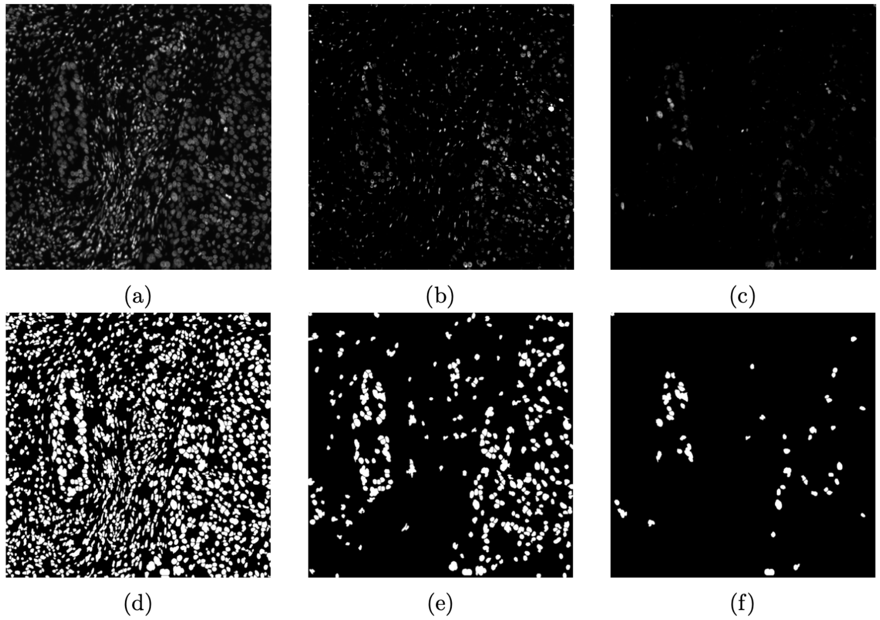


Figure 1: An example set of images with low HR score: a) cells stained with DAPI, b) cells stained with an antibody against geminin protein, c) cells stained with an antibody against RAD51 protein. Using our image processing approach, we d) segmented the cells, e) retained tumor cells positive for geminin, f) retained tumor cells positive for RAD51.

That is,

$$\text{HR Score} = \frac{\text{n (Tumor cells repairing their DNA damage)}}{\text{n (Tumor cells undergoing replication)}}$$

or equivalently,

$$\text{HR Score} = \frac{\text{n (Tumor cells with RAD51 protein)}}{\text{n (Tumor cells with Geminin protein)}}$$

We grouped tumor specimens by their HR scores into two categories: HR-low (0%-35%) and HR-high (>35%). A high HR score indicates a high fraction of tumor cells that are able to repair DNA damage and the patient will possibly not respond to platinum chemotherapy. Similarly, a low HR score means that a low fraction of tumor cells can repair DNA damage and the patient is expected to have a better response platinum chemotherapy. We used an existing dataset of patient's tumorigenic profiles [6] to build our image processing pipeline and to validate HR scores so obtained.

To compute the HR score, we are interested in the tumor cells that are positive for both geminin and RAD51. To count the number of geminin-positive tumor cells, we used the geminin image and compared it with the DAPI image. Next, to count the number of RAD51-positive tumor cells, we used the RAD51 image and compared it with the geminin image. Hence, we filtered out the tumor cells that were positive for geminin only, and those that were positive for RAD51 only. The HR score was calculated using equation 1 discussed earlier.

Even the best algorithms suffer from errors because of under-segmentation and over-segmentation. In order to circumvent this problem, we computed the area of each relevant object and estimated its count proportional to the average area of a tumor cell. For example, if the area of a tumor object was 2.1 times the average area of a tumor cell, we estimated that object to contain two tumor cells. Similarly, if the area of a tumor object was 0.9 times the average area of a tumor cell, we estimated that object to contain one tumor cell. This approach further handled any segmentation errors and improved overall accuracy.

3 RESULTS

We applied the herein implemented pipeline to 45 images from 15 HGSC patients. Figure 1 (a-c) shows images from emission channels for DAPI, Geminin, and RAD51 fluorescent signal. This tumor has low HR capacity, as the number of tumor cells with RAD51 protein (figure 1c) is much lower than the number of tumor cells with geminin positivity (figure 1b).

We then proceed to check whether our pipeline predicts the result accurately. We first pre-processed the DAPI image and segmented the cells using the modified watershed algorithm (figure 1d). Next, we calculated the different object properties to retain geminin-positive tumor cells (figure 1e) using the geminin image (figure 1b), and RAD51-positive tumor cells (figure 1f) using the RAD51 slide (figure 1c), respectively. Finally, we counted the retained tumor cells in figures 1e and 1f and computed the HR score. We obtained a score of 18% (HR-low) against the manually computed score of 19% and correctly established the tumor's HR status.

We further checked if our methodology also worked on tumors with high HR scores. Figure 2 (a-c) shows images from emission channels for DAPI, Geminin, and RAD51 slides of another patient's tumor. This tumor has high HR capacity, as most geminin-positive tumor cells (figure 2b) are also positive for RAD51 protein (figure 2c). We followed the same methodology discussed earlier and got

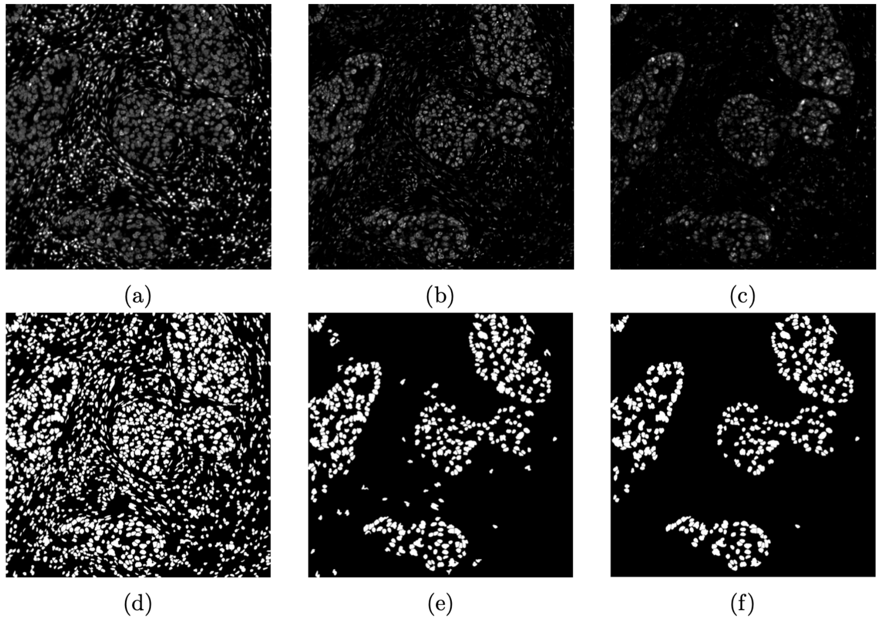


Figure 2: An example set of images with high HR score: a) cells stained with DAPI, b) cells stained with an antibody against geminin, c) cells stained with an antibody against RAD51. Using our image processing approach, we d) segmented the cells, e) retained tumor cells positive for geminin, f) retained tumor cells positive for RAD51.

Table 1: HR scores and their predictions of 15 sets of tumorigenic profiles computed both manually and using our algorithm.

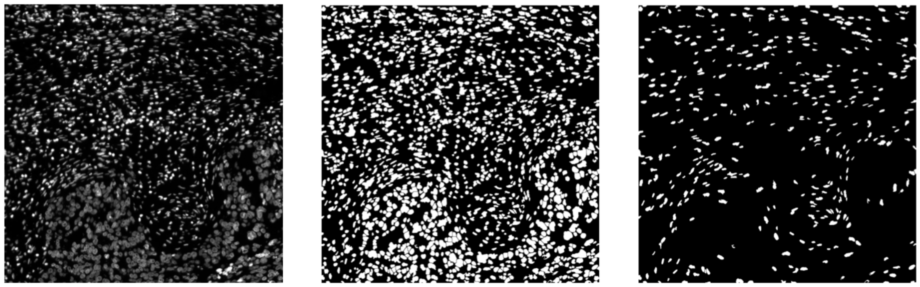
	HR Score			Predicted HR Status	
Set	Manual	Algorithm	Difference (Error)	Manual	Algorithm
1	19.24%	18.09%	1.15%	Low	Low
2	14.65%	16.91%	-2.26%	Low	Low
3	21.92%	18.08%	3.84%	Low	Low
4	25.66%	27.97%	-2.31%	Low	Low
5	22.87%	24.73%	-1.86%	Low	Low
6	58.10%	59.52%	-1.42%	High	High
7	54.11%	59.67%	-5.56%	High	High
8	54.66%	52.08%	2.58%	High	High
9	47.28%	48.68%	-1.40%	High	High
10	79.74%	83.04%	-3.30%	High	High
11	76.57%	81.13%	-4.56%	High	High
12	84.15%	82.32%	1.83%	High	High
13	78.58%	81.10%	-2.52%	High	High
14	75.54%	83.07%	-7.53%	High	High
15	96.26%	98.91%	-2.65%	High	High

a score of 81% (HR-high) against the manually computed score of 80% and thus correctly established the tumor's HR status.

We analyzed images from 15 HGSC patients and computed their HR scores both using a manual approach and the proposed image processing approach. In table 1, we present the HR scores established using both approaches and conclude that our pipeline is able to provide HR scores with high accuracy for each set of images.

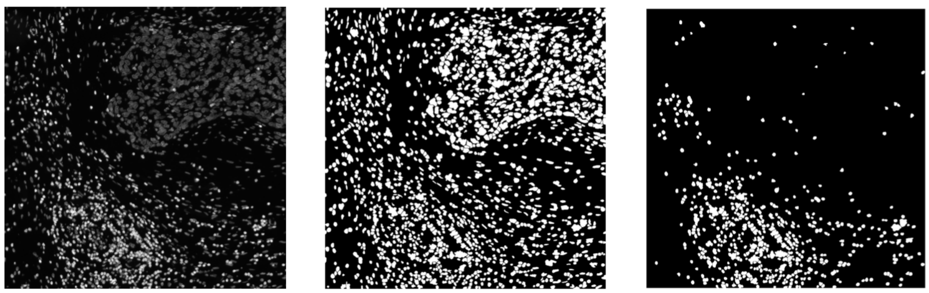
3.1 Quantification of Stromal Cells and Immune cells in Tumor Sections

Our methodology is not just restricted to computing HR scores. We can identify other cell types to analyze, for example, tumor to stromal cells ratio, or tumor to immune cells ratio. In figures 3 and 4, we present the example workings of our pipeline to retain stromal cells and immune cells, respectively. Similar to the above-discussed approach, we have the DAPI images (figures 3a, 4a) and we first



(a) (b) (c)

Figure 3: An example set of images to filter stromal cells: a) cells stained with DAPI, b) segmented image, c) retained stromal cells using object filters.



(a) (b) (c)

Figure 4: An example set of images to filter immune cells: a) cells stained with DAPI, b) segmented image, c) retained immune cells using object filters.

segment them (figures 3b, 4b). We then apply our object filters to retain stromal cells (figure 3c) or immune cells (figure 4c). As shown in Figure 3 and Figure 4, our algorithm is able to distinguish stromal cells and immune cells fairly well and we can further improve its accuracy by estimating the number of cells proportional to the area of each object.

4 CONCLUSION

Machine learning techniques provide means to extract relevant information from medical or biological experiments. With a multitude of side effects associated with platinum chemotherapy, there is a need for tools to predict a patient's response to platinum chemotherapy in advance. Analyses of the homologous recombination pathway and its associated HR scores have established a strong correlation between HR deficiency and predicting long-term survival in ovarian cancer patients. Herein, we proposed an image processing pipeline to compute HR scores using a modified watershed algorithm and filtering cells based on their object properties. Our approach established HR status for 15 HGSC patients. Acquiring more patient data and developing a robust deep learning methodology is a direction for our future work.

ACKNOWLEDGMENTS

This work was supported by the European Union's Horizon 2020 research and innovation programme under Grant Agreement No. 667403 for HERCULES (Comprehensive Characterization and Effective Combinatorial Targeting of High-Grade Serous Ovarian Cancer via Single-Cell Analysis) to S.H, the US National Science Foundation under Grants ECCS-1609236 and ECCS-1917166, the TEES-AgriLife Center for Bioinformatics and Genomic Systems Engineering (CBGSE) startup funds to A.D., the Academy of Finland (grant number 314394 to L.K. and 314395 to S.H.), the Finnish Cancer Foundation and the Sigrid Jusélius Foundation (grants to L.K.).

REFERENCES

[1] Ali, I., A Wani, W., Saleem, K. and Haque, A., 2013. Platinum compounds: a hope for future cancer chemotherapy. Anti-Cancer Agents in Medicinal Chemistry (Formerly Current Medicinal Chemistry-Anti-Cancer Agents), 13(2), pp.296-306.
[2] Fuchs, R.P., 2002. How DNA lesions are turned into mutations within cells?. Oncogene, 21(58).
[3] da Cunha Colombo Bonadio, R.R., Fogace, R.N., Miranda, V.C. and Diz, M.D.P.E., 2018. Homologous recombination deficiency in ovarian cancer: a review of its epidemiology and management. Clinics, 73.
[4] Wiedemeyer, W.R., Beach, J.A. and Karlan, B.Y., 2014. Reversing platinum resistance in high-grade serous ovarian carcinoma: targeting BRCA and the homologous recombination system. Frontiers in oncology, 4, p.34.
[5] Telli, M.L., Timms, K.M., Reid, J., Hennessy, B., Mills, G.B., Jensen, K.C., Szallasi, Z., Barry, W.T., Winer, E.P., Tung, N.M. and Isakoff, S.J., 2016. Homologous recombination deficiency (HRD) score predicts response to platinum-containing neoadjuvant chemotherapy in patients with triple-negative breast cancer. Clinical cancer research, 22(15), pp.3764-3773.
[6] Tumiati, M., Hietanen, S., Hynninen, J., Pietilä, E., Färkkilä, A., Kaipio, K., Roering, P., Huhtinen, K., Alkodsi, A., Li, Y. and Lehtonen, R., 2018. A functional homologous recombination assay predicts primary chemotherapy response and long-term survival in ovarian cancer patients. Clinical Cancer Research, 24(18),

pp.4482-4493.
[7] Luo, J., Wu, M., Gopukumar, D. and Zhao, Y., 2016. Big data application in biomedical research and health care: a literature review. Biomedical informatics insights, 8, pp.BII-S31559.
[8] Dhawan, A.P., 1990. A review on biomedical image processing and future trends. Computer Methods and Programs in Biomedicine, 31(3-4), pp.141-183.
[9] Veta, M., Pluim, J.P., Van Diest, P.J. and Viergever, M.A., 2014. Breast cancer histopathology image analysis: A review. IEEE transactions on biomedical engineering, 61(5), pp.1400-1411.
[10] Adollah, R., Mashor, M.Y., Nasir, N.M., Rosline, H., Mahsin, H. and Adilah, H., 2008. Blood cell image segmentation: a review. In 4th Kuala Lumpur international conference on biomedical engineering 2008 (pp. 141-144)
[11] Irshad, H., Veillard, A., Roux, L. and Racoceanu, D., 2013. Methods for nuclei detection, segmentation, and classification in digital histopathology: a review—current status and future potential. IEEE reviews in biomedical engineering, 7, pp.97-114
[12] Romero-Zaliz, R. and Reinoso-Gordo, J.F., 2018. An updated review on watershed algorithms. Soft computing for sustainability science, pp.235-258.