Bioinformatics, YYYY, 0–0
doi: 10.1093/bioinformatics/xxxxx
Advance Access Publication Date: DD Month YYYY
Applications Note

Bioimage informatics

QuArray: an application for tissue array whole slide image export and signal analysis

Callum Arthurs[1,*], Aamir Ahmed[1]

[1] Prostate Cancer Research Centre at the Centre for Stem Cells and Regenerative Medicine, King's College London, London, United Kingdom

*To whom correspondence should be addressed.

Associate Editor: Jonathan Wren

Received on 18-Oct-2020; revised on XXXXX; accepted on XXXXX

Abstract

Motivation: Tissue array (TA) staining, combined with whole slide imaging (WSI) methods facilitate discovery of biomarkers for diagnosis, prognostication and disease stratification. A key impediment in TA WSI analysis is handling missing tissue and artefacts when identifying tissue cores before quantitative, standardised downstream analysis. There is a need for an open access, user friendly, integrated analysis of the WSI generated using TAs in clinical and scientific research laboratories.

Results: We have developed QuArray (Quantitative Array Application) for image export and signal analysis of TAs using WSI. The application input is a WSI and a corresponding TA configuration file. QuArray identifies and exports core images and analyses chromogen staining in a simple graphical user interface. Output data is saved to file for further analysis including indexed data.

Availability: Available for download from https://github.com/c-arthurs/QuArray under an MIT licence.

Contact: callum.arthurs@kcl.ac.uk

Supplementary information: https://quarray.readthedocs.io

1 Introduction

A tissue array (TA), with hundreds of tissue samples on a single slide, is a valuable biological screening tool used for measurement of RNA and protein expression (Kononen, et al., 1998) allowing reliable and standardised comparison between tissue cohorts. TAs thus facilitate rapid and reproducible analysis of molecular markers and samples.

TA staining and subsequent analysis of digitised whole slide imaging (WSI) files is becoming routine for population-based marker analysis, providing an invaluable resource in cancer biomarker detection (Arthurs, et al., 2017; Botling, et al., 2013; Cao, et al., 2018; Giuliano, et al., 2016; Symes, et al., 2013; Wang, et al., 2010).

The image storage and histological variability in these files make it challenging to create automated WSI analysis pipelines. Ascribing correct tissue co-ordinates can be difficult due to artefacts of TA processing (Kampf, et al., 2012). There are other open source applications for TA analysis, e.g. QuPath (Bankhead, et al., 2017), however, these require scripting knowledge for full automation of image export or analysis from a TA. Our application termed 'QuArray' is an easy-to-use image export and analysis program designed specifically for using chromogen (primarily 3,3'-Diaminobenzidine, DAB) TA, imaged using WSI (TA-WSI).

2 Implementation

We developed the QuArray application to create a platform for TA image export and chromogen analysis in a user friendly, easily accessible and adaptable application that requires no prior programming skills to implement high volume image analysis. QuArray is made to process TA-WSI that have been constructed using methods previously described (Kampf, et al., 2012).

QuArray is written in Python and utilises OpenSlide (Goode, et al., 2013) to interface with the WSI, and the scikit image (Van der Walt, et al., 2014) and NumPy (Van Der Walt, et al., 2011) packages for subsequent image analysis. The user interface was written using the PyQt5 library.

Video files of the sequence of steps described below (and in Figure 1) are provided in the documentation. The program input supports most common WSI formats including SVS (Aperio), SCN (Leica) and NDPI

(Hamamatsu). The user is prompted to create a map of the array as a tabular file (indexed information) where the input fields 0 and 1 can be used to denote missing and present cores, respectively. The user can then implement, either a modifiable tissue core thresholding methodology or manually annotate tissue cores prior to image export. Core images that are exported can be analysed using the chromogen analysis tool after setting a minimum chromogen saturation value in the threshold panel. The output from the application is a set of intensity and coverage results for each image that is matched to indexed tissue information provided by the user.

3 TA core image export

A low magnification RGB image (selected from the image level with a width nearest to 3000 pixels) of the whole slide is loaded into computer memory and displayed to the user. Automatic tissue detection and labelling is achieved by converting the image to greyscale and applying a thresholding method. The pre-set method is Otsu thresholding (Otsu, 1979) but the user can employ others (e.g. Li (Li and Lee, 1993), Triangle (Zack, et al., 1977), or the array Mean). Gaussian blur is applied to the post threshold binary image and the user can select the gaussian sigma value. Binary morphological closing is used to fill gaps that may be present in the tissue so that a core that, for example, is missing or damaged due to sectioning and processing artefacts can be recognised as a single entity for analysis. The slider value for closing will change the size of the selem 2D array that is passed to the closing input; increasing the closing value will result in the union of masked areas that are further apart. A low closing value is advised to reduce CPU computation time. Objects with an area of less than 6000 pixels are generally considered to be noise and are removed from the binary closed mask. The centroid coordinates of each identified object on the WSI are organised into column and row values, accounting for missing cores that were provided in the indexed information. Prior to image export, bounding boxes with labels are presented to the user for confirmation; otherwise, mis-labelled items can be dragged into their correct space. Or the array can be labelled manually by double clicking the centre of each core; the spacebar may be used to undo any erroneous input, serving as an important corrective step.

Images are exported as PNG image files at the highest available magnification level from the WSI (these files are commonly created using a 20x or 40x objective lenses on the slide scanner). The size of the exported image can be allocated prior to export but remains the same for all cores in the TA, for standardisation of analysis.

4 Thresholding and analysis

Images from multiple TA slides can be analysed within the same batch. To do this, chromogen stain analysis is conducted in a separate window. The input of chromogen analysis is a folder of images gathered from the TA export step. QuArray has been optimised for DAB, a commonly used chromogen (secondary label) for protein and single molecule RNA staining. RGB images are loaded into memory and features of the tissue cores are then measured. Total amount of tissue in the image is calculated by measuring the number of pixels that fall above the Triangle (Zack, et al., 1977) threshold value in the greyscale image. The amount of stain present in the core is calculated by separating the image into the HSV colour space and measuring all pixels within the brown hue (a pre-set threshold optimised for DAB staining), above a user defined saturation threshold. The signal per amount of tissue value is calculated by dividing the signal and tissue measurement (Symes, et al., 2013). The mean intensity and standard deviation of stain intensity measurements are also recorded. All results, including user defined indexed information, are saved to a file.

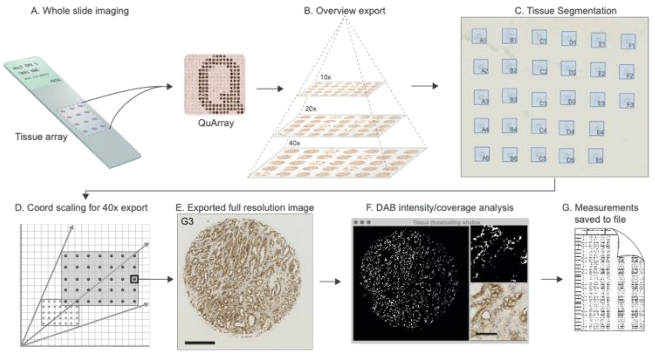

Figure. 1. QuArray application workflow A) The TA is imaged on a slide scanner and the resulting WSI file is fed into QuArray with a user defined map of TA coordinate, indexed with sample information in a tabulated file. B) QuArray is programmed to select a low-resolution image (where the WSI width is nearest to 3000pixels, selected for program speed), for segmentation of tissue areas. This image is loaded into computer memory. C) Automatic or user assisted selection of cores is readied for export. D) TA coordinates are scaled to export the magnified image at the highest resolution available. E) An example of a 1mm diameter DAB labelled human prostate tissue core (scale bar = 250µm) is automatically exported. F) Quantitation of chromogen signal on tissue samples (one core shown as an example). The user is able to select the thresholding parameters based upon chromogen signal saturation. Insets show a whole core image and full resolution (512x512 pixel) zoomed region (scale bar 50µm) of the input image. Chromogen stain at a given threshold is measured. G) All measurements are saved to an excel file for analysis.

5 Conclusions

QuArray is a stand-alone application for exporting images of tissue cores from WSI files for quantification of chromogen stain analysis. QuArray provides an automated, user-friendly application for exporting images acquired from WSI files and measuring signal intensity for biomarker discovery. The application interface is easy to use, requires minimal installation, and is compiled to work on MacOS, Linux, and Windows operating systems.

Acknowledgements

We are grateful to Michael Millar, University of Edinburgh; Dr Christopher Thrasivoulou, University College London; Professor Simon Priestnall, Royal Veterinary College; and Professor Rui Henrique, University of Porto, for help with reagents and expertise. We would also like to thank Bushra Kanwal and Boyu Xie for their help testing the application.

Funding

This work was supported by the Prostate Cancer Research Centre (UK Charity no. 1156027) grant (AA1) to AA. We are also grateful to the King's Development Fund for support.

Conflict of Interest: none declared.

References

Arthurs, C., *et al.* Expression of ribosomal proteins in normal and cancerous human prostate tissue. *PLoS One* 2017;12(10):e0186047.

Bankhead, P., *et al.* QuPath: Open source software for digital pathology image analysis. *Scientific Reports* 2017;7(1):16878.

1 Botling, J., *et al.* Biomarker discovery in non–small cell lung cancer: integrating gene
2 expression profiling, meta-analysis, and tissue microarray validation. *Clinical cancer*
3 *research* 2013;19(1):194-204.
4 Cao, K., *et al.* Quantitative Analysis of Seven New Prostate Cancer Biomarkers and
5 the Potential Future of the 'Biomarker Laboratory'. *Diagnostics (Basel)* 2018;8(3).
6 Giuliano, A., *et al.* Quantitative Expression and Co-Localization of Wnt Signalling
7 Related Proteins in Feline Squamous Cell Carcinoma. *PLoS One*
8 2016;11(8):e0161103.
9 Goode, A., *et al.* OpenSlide: A vendor-neutral software foundation for digital
10 pathology. *Journal of pathology informatics* 2013;4.
11 Kampf, C., *et al.* Production of tissue microarrays, immunohistochemistry staining
12 and digitalization within the human protein atlas. *JoVE (Journal of Visualized*
13 *Experiments)* 2012(63):e3620.
14 Kononen, J., *et al.* Tissue microarrays for high-throughput molecular profiling of
15 tumor specimens. *Nat Med* 1998;4(7):844-847.
16 Li, C.H. and Lee, C.K. Minimum cross entropy thresholding. *Pattern Recognition*
17 1993;26(4):617-625.
18 Otsu, N. A Threshold Selection Method from Gray-Level Histograms. *IEEE*
19 *Transactions on Systems, Man, and Cybernetics* 1979;9(1):62-66.
20 Symes, A.J., *et al.* Quantitative Analysis of BTF3, HINT1, NDRG1 and ODC1
21 Protein Over-Expression in Human Prostate Cancer Tissue. *PLoS. One*
22 2013;8(12):e84295.
23 Van Der Walt, S., Colbert, S.C. and Varoquaux, G. The NumPy array: a structure for
24 efficient numerical computation. *Computing in Science & Engineering*
25 2011;13(2):22.
26 Van der Walt, S., *et al.* scikit-image: image processing in Python. *PeerJ* 2014;2:e453.
27 Wang, Q., *et al.* A Novel Role for Wnt/Ca2+ Signaling in Actin Cytoskeleton
28 Remodeling and Cell Motility in Prostate Cancer. *PLOS ONE* 2010;5(5):e10456.
29 Zack, G.W., Rogers, W.E. and Latt, S.A. Automatic measurement of sister chromatid
30 exchange frequency. *J Histochem Cytochem* 1977;25(7):741-753.

31