Declan C. Murphy, MBBS, MRes[1,2]
Amar V. Nasrulloh, BSc[3,4]
Clare Lendrem, PhD, CStat[5]
Sara Graziado, PhD, MPhil[6]
Mark Alberti, BSc[7]
Morten la Cour, PhD[7]
Boguslaw Obara, PhD[3]
David H.W. Steel, MBBS, FRCOphth, MD[1,8]

[1]*Newcastle University, Newcastle Upon Tyne, United Kingdom;* [2]*Northumbria Healthcare NHS Foundation Trust, Tyne and Wear, United Kingdom;* [3]*Department of Computer Science, Durham University, Durham, United Kingdom;* [4]*Physics Study Program, University of Lambung Mangkurat, Banjarbaru, Indonesia;* [5]*National Institute for Health Research Newcastle In Vitro Diagnostics Co-operative, Newcastle University, Newcastle Upon Tyne, United Kingdom;* [6]*National Institute for Health Research Newcastle In Vitro Diagnostics Co-operative, Newcastle upon Tyne Hospitals Foundation Trust, Newcastle Upon Tyne, United Kingdom;* [7]*Department of Ophthalmology, Rigshospitalet, University of Copenhagen, Copenhagen, Denmark;* [8]*Sunderland Eye Infirmary, Sunderland, United Kingdom*

The study was supported by grants from the British and Eire Association of Vitreoretinal Surgeons (BH190620) and the Newcastle University & Wellcome Trust institutional strategic support fund (BH184036). This project is supported by the National Institute for Health Research (Newcastle In Vitro Diagnostics Co-operative). The views expressed are those of the author(s) and not necessarily those of the National Institute for Health Research or the Department of Health and Social Care. Financial Disclosure(s): The author(s) have made the following disclosure(s): D.H.W.S.: Consultant – Alcon, Roche, Gyroscope; Research funding – Alcon, Bayer, all unrelated to this study. A.V.N. was sponsored by the Indonesian Endowment Fund for Education, Indonesia (LPDP Indonesia) No. 20150422012873. No other authors have any financial disclosures to make.

HUMAN SUBJECTS: Data were obtained from a previously published randomized controlled trial that obtained ethical approval (protocol Number: H-4-2013-091, Rigshospitalet, Copenhagen) and full informed consent from all participants. All research adhered to the tenets of the Declaration of Helsinki.

No animal subjects were used in this study.

Author Contributions:
Conception and design: Murphy, Lendrem, Graziado, Obara, Steel
Data collection: Murphy, Nasrulloh, Alberti, la Cour, Obara, Steel
Analysis and interpretation: Murphy, Lendrem, Graziado, Steel
Obtained funding: Steel
Overall responsibility: Murphy, Nasrulloh, Lendrem, Graziado, Alberti, la Cour, Obara, Steel

Abbreviations and Acronyms:
AL = axial length; **BD** = base diameter; **MH** = macular hole; **MLD** = minimum linear diameter; **SD-OCT** = spectral domain OCT; **3D** = 3-dimensional; **VA** = visual acuity.

Correspondence:
David H.W. Steel, MBBS, FRCOphth, MD, Sunderland Eye Infirmary, Queen Alexandra Road, Sunderland, UK. E-mail: david.steel@ncl.ac.uk.

References

1. Steel DH, Donachie PHJ, Aylward GW, et al. Factors affecting anatomical and visual outcome after macular hole surgery: findings from a large prospective UK cohort. *Eye (Lond)*. 2020 Mar 30 [Epub ahead of print].
2. Duker JS, Kaiser PK, Binder S, et al. The International Vitreomacular Traction Study Group Classification of Vitreomacular Adhesion, Traction, and Macular Hole. *Ophthalmology*. 2013;120:2611–2619.
3. Chen Y, Nasrulloh AV, Wilson IH, et al. Macular hole morphology and measurement using an automated three dimensional image segmentation algorithm. *BMJ Open Ophthalmol*. 2020. in press.
4. Geng XY, Wu HQ, Jiang JH, et al. Area and volume ratios for prediction of visual outcome in idiopathic macular hole. *Int J Ophthalmol*. 2017;10:1255–1260.
5. Xu D, Yuan A, Kaiser PK, et al. A novel segmentation algorithm for volumetric analysis of macular hole boundaries identified with optical coherence tomography. *Investig Ophthalmol Vis Sci*. 2013;54:163–169.
6. Nasrulloh AV, Willcocks CG, Jackson PTG, et al. Multi-Scale Segmentation and Surface Fitting for Measuring 3-D Macular Holes. *IEEE Trans Med Imaging*. 2018;37:580–589.
7. Alberti M, la Cour M. Nonsupine positioning in macular hole surgery: a non-inferiority randomized clinical trial. *Retina*. 2016;36:2072–2079.

Gene Expression Profile Prediction in Uveal Melanoma Using Deep Learning

A Pilot Study for the Development of an Alternative Survival Prediction Tool

In recent years, artificial intelligence, especially deep learning (DL), has generated immense interest in the medical field. Deep learning has been used to classify medical images in disciplines such as ophthalmology and oncologic pathology. One commonality across malignancies is that cancer cell morphologic features potentially reflect the underlying genetics and that careful analysis of cytopathologic characteristics often provides helpful prognostication information. However, detailed measurement and analysis of cell morphologic features are labor intensive and clinically infeasible, and thus is limited largely to research. Analyses of pathologic images to extract useful information ultimately are a pattern recognition exercise in which DL excels. We hypothesize that DL methods, when applied appropriately in cytopathologic image analysis, could predict patient outcomes that correlate with the tumors' genetic or molecular profiles, or both. Our disease of interest is uveal melanoma (UM), which is unique among malignancies for having a validated prognostic gene expression profile (GEP) test that can be used independently of other clinicopathologic parameters and can be tested on fine-needle aspiration biopsy (FNAB) samples. Patients with UM can be divided into 2 classes by GEP, with a survival probability of 95% in class 1 patients and 31% in class 2 patients at 92 months.[1,2] Our ultimate goal is to develop a DL-based image analytic tool for survival prognostication in UM. Given that GEP is correlated highly with survival in UM, we set out to conduct a pilot study to develop a DL system that can distinguish patient survival using smeared cytologic aspirates from FNAB samples and GEP as the reference standard.

Our retrospective study was conducted in accordance with the tenets of the Declaration of Helsinki and was approved by the institutional review board. No informed consent was obtained as this was a retrospective study. The Johns Hopkins University IRB reviewed this study and determined it to be exempt. In total,

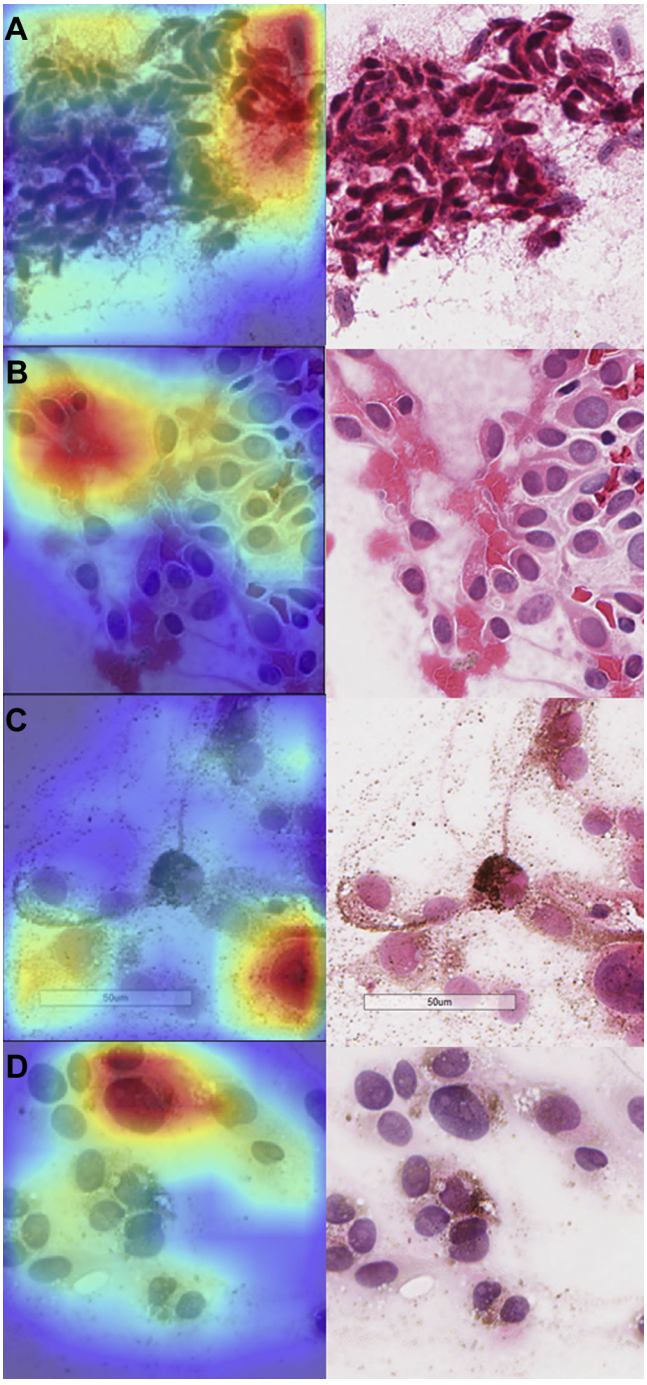

Figure 1. Sample class activation mapping analyses of correctly predicted cytopathologic image tiles. **A**, Patient 5, gene expression profile (GEP) class 1. The highlighted cells demonstrate classic spindle morphologic features, which are associated with better prognosis. **B**, Patient 10, GEP class 1. The highlighted cells exhibit less atypia than the rest of the cells. **C**, Patient 13, GEP class 2. The highlighted cell exhibits epithelioid cytomorphologic features, which are known to carry a worse prognosis. **D**, Patient 18, GEP class 2. The highlighted region contains a cell with the highest nuclear-to-cytoplasmic ratio and degree of atypia, features that are associated with a worse prognosis.

20 de-identified, FNAB cytologic slides stained with hematoxylin–eosin were selected randomly (1 slide per patient; 10 from GEP class 1 and 10 from GEP class 2) from one of the author's (Z.M.C.) cohort of UM patients.[3] Whole-slide scanning was performed. Using a magnification of ×40, native-resolution crops containing UM cells were saved in TIFF format and were split further into 8 tiles of equal size. The tiles then were selected for further processing only if at least 1 UM cell was present. Typically, each slide generated hundreds of ×40 snapshot images, and of the 20 slides, a total of 26 351 unique tiles were generated.

Transfer learning and the ResNet-152[4] deep convolutional neural network were adopted, and the last fully connected layer was modified for our binary classification problem of distinguishing between patients with good and poor prognosis per GEP classification. We performed leave-1-out cross-validations. To test each of the 20 slides (and patients), we trained 10 models with a different training and validation split, expecting our DL system's performance to be affected by the data split because of the small dataset size and patient-level splitting. Specifically, for each of the leave-1-out cross-validations, we performed 10 random samplings for the validation subset selection. If slide 1 was used as the testing slide, then the other 19 slides were used for model development: 17 slides for training and 2 slides for validation (1 from class 1 and 1 from class 2). Slide 1 then was tested 10 different times by 10 different models that were generated by 10 random and different combinations of training and validation slides. For example, model 1 would use slide 2 and slide 11 for validation. Model 2 would use slide 3 and slide 12 for validation, and so forth. Eventually, 10 models were generated, and the mean accuracy of these 10 models was obtained. If the lower 95 confidence interval value exceeded 50%, then we concluded that the GEP of slide 1 (i.e., patient 1) was predicted correctly. This process was repeated for all 20 slides (and patients), such that each slide (or patient) was evaluated 10 times by 10 different models. To identify the image features used by the deep convolutional neural network to predict GEP, we created heatmaps through class activation mapping,[5] a technique that visually highlights areas of importance for classification decisions within an image (the warmer the color—e.g., red—the higher the importance of that region).

In our pilot study, we were able to correlate the DL prediction with GEP in 15 of 20 patients from our cohort (point estimate of 75% accuracy; 95% confidence interval, 51%–91%). Given that GEP is correlated highly with survival, the data suggest that prognostication information can be predicted from hematoxylin–eosin pathology slides alone in UM using DL. Sample class activation mapping analyses (Fig 1) for the correctly predicted images showed that our deep convolutional neural network generally focused on biologically relevant features: UM cells with spindle-shaped morphologic features or less atypia in GEP class 1 images and UM cells with epithelioid morphologic features, more atypia, and larger nuclei or nucleoli in GEP class 2 images. Interestingly, our algorithm predicted "poor outcome" in a patient with a class 1 tumor who died of unexpected early metastasis 28 months after the initial diagnosis, and predicted a "not so unfavorable outcome" in 2 class 2 patients who survived for more than 20 months after metastasis was detected, significantly longer than the median survival time of 3.9 months.[6] If this trend can be reproduced prospectively and validated externally with actual survival data as the reference standard, it would suggest that a mature version of our algorithm may provide

more fine-grained survival prognostication when paired with GEP testing: predict unfavorable clinical surprises in class 1 patients and extended survival in class 2 patients.

Our pilot study has numerous limitations. First, FNABs can be technically challenging and susceptible to sampling errors, although all biopsies were performed by the same experienced ocular oncologist (Z.M.C.). Second, currently available saliency analysis techniques, such as class activation mapping, are only partially explainable. For example, it is unclear how our algorithm decides which cells to focus on to make predictions. Third, although our DL system was developed with more than 25 000 unique data points, it ultimately included data from only 20 UM patients. The small patient sample size and data variation required us to perform leave-1-out validations, instead of the more conventional method of training and validating a single model. As the next step, we plan to train an algorithm with a larger database and actual survival data as the reference standard and to test it against an external dataset, with the goal of developing an alternative survival prediction tool in UM.

T. Y. Alvin Liu, MD[1]
Hongxi Zhu, BS, MS[2]
Haomin Chen, BS, MA[3]
J. Fernando Arevalo, MD, PhD[1]
Ferdinand K. Hui, MD[4]
Paul H. Yi, MD[4]
Jinchi Wei, BSE[5]
Mathias Unberath, PhD[6]
Zelia M. Correa, MD, PhD[1,7]

[1]*Wilmer Eye Institute, Johns Hopkins University, Baltimore, Maryland;* [2]*Computational Interaction and Robotics Lab, Johns Hopkins University, Baltimore, Maryland;* [3]*Department of Computer Science, Johns Hopkins University, Baltimore, Maryland;* [4]*Department of Radiology, Johns Hopkins University, Baltimore, Maryland;* [5]*Department of Biomedical Engineering, Johns Hopkins University, Baltimore, Maryland;* [6]*Malone Center for Engineering in Healthcare, Johns Hopkins University, Baltimore, Maryland;* [7]*Department of Ophthalmology, University of Cincinnati College of Medicine, Cincinnati, Ohio*

Financial Disclosure(s): The author(s) have made the following disclosure(s): J.F.A.: Consultant - DORC International B.V., Allergan, Inc., Bayer, Mallinckrodt; Financial support - Topcon; Royalties - Springer SBM LLC.

Z.M.C.: Financial support - Castle Biosciences, Inc., Immunocore, LLC.

Supported by the Emerson Collective Cancer Research Fund, Palo Alto, CA (grant no.: 642653 [T.Y.A.L.]; the Unrestricted Grant to Wilmer from Research to Prevent Blindness, New York, NY (M.U.); the Radiological Society of North America (P.H.Y.).

HUMAN SUBJECTS: Human subjects were included in this study. The IRB at Johns Hopkins University approved the study. All research adhered to the tenets of the Declaration of Helsinki. No informed consent was obtained as this is a retrospective study.

No animal subjects were included in this study.
Author Contributions:
Conception and design: Liu, Correa
Analysis and interpretation: Liu, Zhu, Chen, Arevalo, Hui, Yi, Wei, Unberath, Correa
Data collection: Liu, Zhu, Correa
Obtained funding: Liu
Overall responsibility: Liu, Zhu, Chen, Arevalo, Hui, Yi, Wei, Unberath, Correa

Abbreviations and Acronyms:
DL = deep learning; **FNAB** = fine-needle aspiration biopsy; **GEP** = gene expression profile; **UM** = uveal melanoma.

Correspondence:
T. Y. Alvin Liu, MD, Wilmer Eye Institute, Johns Hopkins University, 600 North Wolfe Street, Maumenee 726, Baltimore, MD 21287. E-mail: tliu25@jhmi.edu.

References

1. Onken MD, Worley LA, Ehlers JP, Harbour JW. Gene expression profiling in uveal melanoma reveals two molecular classes and predicts metastatic death. *Cancer Res*. 2004;64:7205–7209.
2. Onken MD, Worley LA, Char DH, et al. Collaborative ocular oncology group report number 1: prospective validation of a multi-gene prognostic assay in uveal melanoma. *Ophthalmology*. 2012;119:1596–1603.
3. Correa ZM, Augsburger JJ. Independent prognostic significance of gene expression profile class and largest basal diameter of posterior uveal melanomas. *Am J Ophthalmol*. 2016;162:20–27 e21.
4. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016:770–778.
5. Zhou B, Khosla A, Lapedriza A, et al. *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016: 2921–2929.
6. Lane AM, Kim IK, Gragoudas ES. Survival rates in patients after treatment for metastasis from uveal melanoma. *JAMA Ophthalmol*. 2018;136:981–986.