Patterns

CellPress
OPEN ACCESS

People of Data

Oner, Sung, and Lee: Researchers in digital pathology for the future of modern medicine

Mustafa Umit Oner,[1,2] Wing-Kin Sung,[2,3,*] and Hwee Kuan Lee[1,2,4,5,6,7]
[1]Bioinformatics Institute, Agency for Science, Technology and Research (A*STAR), Singapore 138671, Singapore
[2]School of Computing, National University of Singapore, Singapore 117417, Singapore
[3]Genome Institute of Singapore, A*STAR, Singapore 138672, Singapore
[4]Singapore Eye Research Institute (SERI), Singapore 169856, Singapore
[5]Image and Pervasive Access Lab (IPAL), Singapore 138632, Singapore
[6]Rehabilitation Research Institute of Singapore, Singapore 308232, Singapore
[7]Singapore Institute for Clinical Sciences, Singapore 117609, Singapore
*Correspondence: ksung@comp.nus.edu.sg
https://doi.org/10.1016/j.patter.2022.100447

Oner, an early-career researcher, and Lee and Sung, group leaders, have developed a deep learning model for accurate prediction of the proportion of cancer cells within tumor tissue. This is a necessary step for precision oncology and target therapy in cancer. They talk about their view of data science and the evolution of pathology in the coming years.

Hwee Kuan, Mustafa, and Wing-kin, tell us about your background (personal and/or professional)

Hwee Kuan Lee: My background training is theoretical physics, but I switched fields into bioinformatics 15 years ago. My current research interests are foundational AI algorithm development, AI applications in clinical research, AI applications in physics, and biomolecular simulations.

Mustafa Umit Oner: I got my BS (2013) and MS (2016) degrees from the Electrical and Electronics Engineering Department at the Middle East Technical University in Turkey. I worked on deep learning and its application in digital histopathology during my masters. The deep-learning-based algorithm that I developed made us the fourth-ranked team in the CAMELYON16 ISBI challenge on cancer metastases detection in lymph nodes, the first challenge ever using whole-slide images. Then, I decided to pursue a PhD in medical AI and came to Singapore, one of the best places in the world to study interdisciplinary research in this domain, I think. I was lucky to have two excellent supervisors, Wing-Kin Sung and Hwee Kuan Lee, providing me with precious mentoring and the resources of productive working environment, a robust research infrastructure, and great collaborations. More importantly, I am grateful for the independence they provided me to explore my way.

Wing-Kin Sung: I received both the BSc and PhD in the Department of Computer Science from the University of Hong Kong in 1993 and 1998, respectively.

Hwee Kuan, tell us about your journey to your current position, the application process for your position, and setting up your own team. What were the challenges? What institutional support has been particularly important?

HKL: I became a principal investigator in 2006 when the institute started a new research division. I was a collaborator for the previous institute director, and he appreciated my skill set.

Support from collaborators within the institute as well as support from admin/IT team has been important for establishing my own team.

Hwee Kuan, how was the experience publishing your first independent paper, the first paper from your own team? How did it differ from publishing as first author on your previous papers?

HKL: The experience was good, but I don't feel much of a difference from publications when I was a postdoc. Essentially good science that helps humankind is good science; does not matter who publishes it.

Hwee Kuan and Wing-Kin, tell us about the research in your team. What drew you to this area of research? How has the research focus of your team evolved over the years?

HKL: We started with focusing on analysis of cellular images. Now we also work on radiology images, biomolecular simulations, and agriculture technologies

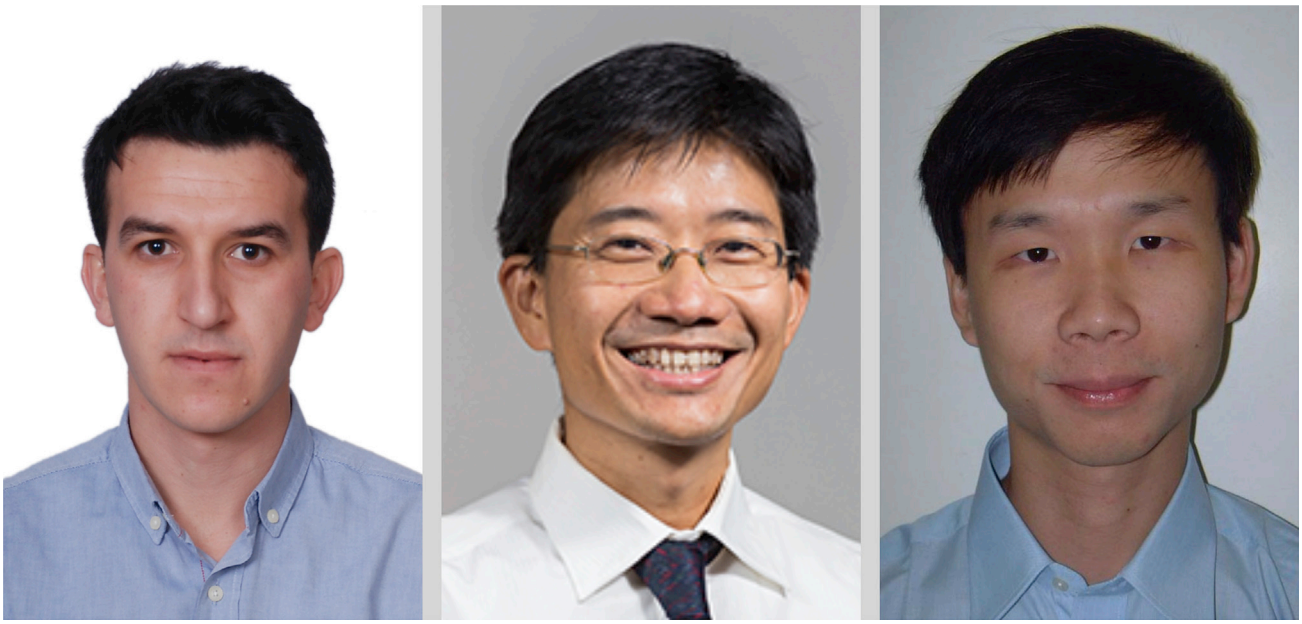
W-KS: My team is focused on genomics. We currently are focused on human and plant research.

Hwee Kuan, where is the team based currently? How long have you been there? Who currently makes up the team? Tell us a bit about them

HKL: The team is in Singapore. I have been in this research institute for 15 years. The laboratory consists of five postdocs, five PhD students, and one technician. They are all working on AI-related research.

Mustafa, what drew you to your current team and topic?

MUO: After my master's, I looked for an interdisciplinary group working on medical AI in general and digital pathology in particular. I joined the group in 2017 as a PhD student to explore the potential of deep learning in digital pathology with a team of experts from different fields, like pathology, oncology, genomics, biology, and machine learning. We always



From left to right: Mustafa Umit Oner, Wing-Kin Sung, and Hwee Kuan Lee

consider the clinical value of a work. High-throughput genomic analysis on sequencing data has become an indispensable tool for cancer research and has enabled precision oncology. However, accurate tumor purity prediction while selecting samples for sequencing was a challenge. Therefore, we focused on this task, which has been recently published in *Patterns*.[1]

Hwee Kuan and Wing-Kin, what kind of atmosphere do you look to foster in your team?

HKL: An atmosphere where everyone realizes that by sharing and teaching one another, everyone benefits. We also need to be very critical of our colleagues' work.

W-KS: I hope to have an environment that encourages the team members to discuss. Without discussion, it is difficult to generate new ideas.

Hwee Kuan, aside from supervising their research, how do you help to develop and mentor your students and postdocs as data scientists? How do you adapt your mentoring style to different people?

HKL: Technical skills are secondary to soft skills such as communication skills, sharing, and caring for a collaborator.

I assign projects according to interest, motivation, and skillset. Some people need clear instructions. Some need more freedom.

Hwee Kuan, Mustafa, and Wing-Kin, what motivated you to become data researchers? Is there anyone/anything in particular that helped guide you on your path?

HKL: When I started my work on bioinformatics, it soon became obvious that the practical requirements in bioinformatics and translational clinical research calls for machine learning and data-driven methods. Domain knowledge is important, but there is no theoretical biology like those in theoretical physics. Therefore, there are no good models like those in physics to do predictions. I had been helped by many colleagues, including clinicians, biologists, and fellow data scientists.

MUO: My primary motivation is learning new things and the peculiar feeling that is a mixture of excitement, satisfaction, and bewilderment at the time of learning or understanding something. I guess that feeling brought me up to today with a lot of fun and, of course, failure and struggle on the way. Still, that feeling keeps me on track during my journey of trying to be a good researcher. Besides, I always try to do my best and keep my eyes open for opportunities to improve. I think being a good researcher is a form of evolution.

W-KS: There are a lot of hidden jewels in the data. As a computer scientist, I hope to discover some interesting and useful results. In particular, I am interested in questions related to health. Hence, I started to work in bioinformatics, a branch in data science.

Hwee Kuan, Mustafa, and Wing-Kin, what is the definition of data science in your opinion? What is a data scientist? Do you self-identify as one?

HKL: In my opinion, data science is a field of science to perform various regression using data-driven approaches.

MUO: Data science is the collection of all activities systematically conducted over data to reveal patterns engraved within the data, which eventually helps us understand the system generating the data. It is not an isolated step of processing data with a simple input-output approach. On the contrary, it requires a systems thinking approach incorporating various disciplines' contributions at different stages. Then, a data scientist is the person successfully collaborating with the experts and translating their domain knowledge into the process of

data science with the help of data manipulation and critical-thinking skills to understand the system.

W-KS: To me, data science is the field to understand the real world through analysis of the digital data observed from the real world.

Hwee Kuan and Wing-Kin, tell us about any barriers you faced in pursuing data science as a career

HKL: It is tempting to just feed in data and look for results, because this is the most straightforward and "lazy" way of drawing conclusions. However, we know very well that we need to use domain knowledge to understand and interpret our conclusions.

W-KS: Collecting suitable real data is not easy.

Mustafa and Wing-Kin, which of the current trends in data science seem the most interesting to you? In your opinion, what are the most pressing questions for the data science community?

MUO: The immense potential of data science due to its interdisciplinary nature makes it attractive to me. As researchers from various domains find better ways of effectively working together, this potential will transform our lives. Indeed, how we can effectively work together and melt our knowledge in the same pot is one of the most pressing questions in data science for me.

W-KS: I am currently interested in bioinformatics. The current key issue is how can the community share data.

Hwee Kuan and Mustafa, what is the role of data science in the domain/field that you work in? What advancements do you expect in data science in this field over the next 2–3 years?

HKL: I work in theoretical AI development and AI applications in biomedical domains. In the next 2–3 years, AI applications in biomedical domain will mature, and we will see more FDA-approved AI methods.

MUO: Data science is transforming the clinical workflow in pathology. Recently, an AI product in digital pathology got the first-ever FDA approval for prostate cancer detection.[2] These developments will ease the lives of medical professionals, and they can also improve patient outcomes. In the coming years, probably, we are going to see more of these products. Different than that, I expect to see more studies producing scientific discoveries by integrating interdisciplinary data and knowledge into the data science workflow.

Hwee Kuan, Mustafa, and Wing-kin, how do you keep up to date with both advances in data science techniques and advances in the field/domain that you work in?

HKL: Serve as reviewers for journals and conferences, do journal clubs, systematic literature reviews for every scientific project we are doing.

MUO: I follow publications in high-quality journals and conferences. I also follow the social media accounts and email groups of research teams in the related fields, which is a fast and efficient way of learning about recent advancements and being aware of early archival versions of the studies.

W-KS: Learn from collaborators and read papers.

Hwee Kuan, which achievement/discovery in your career are you most proud of? Tell us about it

HKL: Deriving a mathematical framework for time travel and showed that paradoxes can be precluded.[3]

Hwee Kuan, a lot of data scientists continue their career outside of academia. Do you encourage your students and postdocs to continue their career in academia and/or outside of academia?

HKL: Yes, I am supportive of careers within and outside of academia. Different people are suited for different jobs

Mustafa, what's next for the project? And finally, what's next for you?

MUO: The next step for the project is validating our findings with multi-region sequencing and applying the developed model to other clinically relevant features such as immune cell infiltration prediction, gene mutation prediction, and intra-tumor heterogeneity prediction. For me, I am looking for my next endeavors as an independent researcher.

Mustafa, why did your team decide to publish in *Patterns*?

MUO: *Patterns* has a broad readership interested in cross-disciplinary data science problems. Its high-quality publishing values ensure bringing groundbreaking data science research to the correct audience. We thought *Patterns* is a perfect venue for our study that discusses the challenges and importance of comprehensive tumor purity analysis in machine learning and medical domains.[1]

Hwee Kuan, what is your advice for future data scientists?

HKL: Always understand the domain knowledge and build this knowledge into your data science workflow.

REFERENCES

1. Oner, M.U., Chen, J., Revkov, E., James, A., Heng, S.Y., Kaya, A.N., Alvarez, J.J.S., Takano, A., Cheng, X.M., Lim, T.K.H., et al. (2021). Obtaining spatially resolved tumor purity maps using deep multiple instance learning in a pan-cancer study. Patterns *3*, 100399.
2. U.S. Food & Drug Administration (2021). FDA Authorizes Software that Can Help Identify Prostate Cancer. News release from the U.S. Food & Drug Administration, September 21, 2021. https://www.fda.gov/news-events/press-announcements/fda-authorizes-software-can-help-identify-prostate-cancer.
3. Lee, H.K. (2011). Transition Matrices and Time Travel. Phys. Procedia *15*, 59–63.

About the authors

Mustafa Umit Oner got his BS and MS degrees from the Electrical and Electronics Engineering Department at the Middle East Technical University, Ankara, Turkey, in 2013 and 2016, respectively. Then, he got his PhD degree in 2021 from the Computer Science Department at the National University of Singapore, Singapore. He continues his research at the Bahcesehir University, Istanbul, Turkey. He is interested in developing novel machine learning models and machine-learning-based information systems for digital histopathology and integrative multi-omics to support diagnostic and therapeutic decision making in cancer.

Wing-Kin Sung received both a BSc and a PhD in the Department of Computer Science from the University of Hong Kong in 1993, 1998, respectively. He is a professor in the Department of Computer Science, School of Computing, NUS, and a senior group leader in Genome Institute of Singapore. He has over 25 years of experience in algorithm and bioinformatics research. He has published over 230 high-impact papers. He was conferred the 2003 FIT paper award (Japan), the 2006 National Science Award (Singapore), and the 2008 Young Researcher Award (NUS) for his research contribution in algorithm and bioinformatics.

Hwee Kuan Lee obtained his PhD in 2001 in theoretical physics from Carnegie Mellon University. He

is currently a senior principal investigator and the head of the Cellular Image Informatics Division at the Bioinformatics Institute, Agency for Science, Technology and Research, Singapore. At the same time, he is the deputy director for training and talent. He has also been appointed joint and adjunct positions in several other institutions, such as National University of Singapore, Nanyang Technological University, Singapore Institute for Clinical Sciences. His research interests are in theoretical AI development and their applications in the biomedical field.