

NII Shonan Meeting Report

No. 2018-17

Patient Similitude: Combining Histopathological Images & Multiple-Scale Molecular Phenotypes

Raghu Machiraju
Motonori Ota
Jens Rittscher

November 12-15, 2018



National Institute of Informatics
2-1-2 Hitotsubashi, Chiyoda-Ku, Tokyo, Japan

Patient Similitude: Combining Histopathological Images & Multiple-Scale Molecular Phenotypes

Contents

1	Foreword	2
2	The Premise	3
3	The Rationale	3
4	Goals of the workshop	4
5	Meeting schedule	4
6	Organisers	6
7	Talk abstracts	8
7.1	Junya Fukuoka	8
7.2	Andrey Bychkov	9
7.3	Kishio Kuroda	10
7.4	Robert West	11
7.5	Carolina Wälby	12
7.6	Gianni Monaco	13
7.7	Lee Cooper	14
7.8	Jun Xu	15
7.9	Korsuk Sirinukunwattana	16
7.10	Kazuaki Nakane	17
7.11	Guadalupe Canahuate	18
7.12	Hirohisa Oda	19
7.13	Seiichi Uchida	20
7.14	Saadia Iftikhar	21
7.15	Discussion: AI Expectations & Challenges	22
7.16	Liz Marai	23
7.17	Mitsuyuki Nakao	24
7.18	David Wedge	25
7.19	Anil Parwani	26
8	Challenge Discussion	27



Seminar



Mount Fuji (finally!)

Figure 1: **Photos.** *In addition to the very official group picture we also wanted to include some evidence that we accomplished a lot. On the last day we had a clear view of Mount Fuji.*

9 Workshop Outcomes	27
9.1 Research collaborations initiated by participants	28
9.2 New research questions	28
9.3 Activities initiated by the entire group	28
10 List of participants	29

1 Foreword

The organisers would like to express special *Thank You!* to the Scientific Committee of the Shonan Conference Centre for enabling this scientific meeting. As a very diverse group that included pathologists, geneticists, engineers and computer science researchers we enjoyed a very unique opportunity of discussing the topic of utilising histopathology in the clinical setting.

The group identified a number of opportunities for future collaborative research. Central to these discussion was the planning of a an international challenge contest, which is captured in Section 8.

Due to health reasons one of the organizers, Prof. Raghu Machiraju, was not able to attend the meeting. We would like to express special recognition to Prof. Elisabeta (Liz) Marai who volunteered to coordinate the meeting. Her interactive and stimulating brainstorming sessions helped lay the foundations for our discussions on the last day.

We truly enjoyed the Japanese hospitality. The group will treasure the memories of the trip to Kamakura and the traditional Japanese dinner. On the very last day we got a clear view of Mount Fuji - undoubtedly one of the highlights of our visit.

2 The Premise

Patient data is increasingly available in many forms including both genomic and various phenotypic forms including - transcriptomic, epigenetic, proteomic, histopathological, radiological, and clinical variables. Repositories such as The Cancer Genome Atlas (TCGA) contain multitude of patient records which can be used for patient stratification, disease subtype/biomarker discoveries, all essential elements in the era of personalized medicine. While there is the promise that the integration of all this available information will help to realise a more personalised and precise treatment, salient fundamental and foundational questions need to be addressed before this vision can become a reality.

Motivating example: Two patients afflicted with a Stage 3 of a certain cancer can report a similar gene panel yet the ensuing analysis can still lead to different survival regimens since the expressed regulatory microRNA expression are different. The outcome to treatment and hence the composite response is indeed very different. The need is even more urgent given the requirement to assess outcome to treatment. One example for such therapeutic advances is cancer immunotherapy. While this new form of treatment extends life expectancy by several years and not just weeks, we still do not understand which patients respond to treatment and which do not.

3 The Rationale

Below, we first provide a sound rationale for our approach especially for the comprehensive use of both genomic and phenotypic information including histology images. **A case for integrative methods:** Fundamental methodological advances are necessary to achieve this goal. While genomic information provides important *a priori* information on potential risk profiles, gene expression and transcript data tells us which genes are activated to likely produce salient proteins, metabolomes and other phenotypic traits. Epigenetic, transcriptional regulation, and post-translational modifications can manifest in specific patterns at the tissue scale and often captured as histology images. Studies on large patient cohorts, such as genome wide association studies, can only provide partial answers. The stratification of patients into subgroups that not only have a similar genomic risk profiles but also show the same therapeutic response is a question we cannot answer with existing approaches and without considering especially histopathological information. Although direct modeling methodologies exist across the scales, machine learning methodologies rarely transcend data across scales. It is the goal of this workshop to begin addressing this topic in a rigorous fashion.

Role of histopathology images: Still quantitative data derived from images taken at the cellular and tissue level has not played the role that it potentially can in assisting clinical decision making. We submit that the association of image based phenotypes with genotypes is also necessary for identifying disease subtypes that are consistent with our molecular understanding of disease. Imaging provides important contextual and spatial information on multiple different anatomical levels. Thanks to advances in medical image analysis, computer vision and machine learning we have made tremendous progress in analyzing medical images acquired on CT, MR, PET or ultrasound machines. However, histopathology images have not yet benefited from a systemic studies given their large size and visual complexity. Enabled by

the more widespread availability of digital slide scanners, the field of computational and digital pathology is now one of the challenge areas of biomedical image analysis. Further, the advances in deep learning and methods for processing complex images are encouraging a closer scrutiny of this collection of data with new approaches and methodologies.

Broader Questions and challenges: Evidence-based methods that point to a multidimensional and multifactorial definition of patient similarity or similitude is certainly the need of the hour. Today, most clinical decision making relies on categorical grading systems that do not reflect our current understanding of disease. This current practice enforces categorical boundaries of variables that are in fact continuous and quantitative. As argued earlier, we believe that the concept of patient similarity is key to the discovery of new disease-relevant subtypes and effective patient management.

4 Goals of the workshop

In this workshop we addressed methodological and application specific challenges of a principled data integration in the genomic and phenotypic spaces. Such an integration will allow for the differential comparison of patients and their outcomes. As histopathology plays a significant role in clinical decision making and patient management, they serve as the linchpin of integrative methods. For instance, the National Pathology Programme in the UK estimates that pathology is involved in 70% of all diagnoses offered by Britain's National Health Service. We explored approaches in the workshop that built on the existing diagnostic workflows but extend them to integrate other available information. The conducted workshop provided an unique interdisciplinary setting and brought together clinical scientists, bioinformaticians as well as computer scientists and biomedical engineers to discuss these challenges and nucleate a scientific community.

The Pertinent Questions: The workshop aimed to pose these three primary questions:

- How can we make the transition from categorical disease descriptions to representations that will foster the development of evidence-based notion of patient similarity?
- How does one formulate composite measures of similarity that work across the genomic and phenotypic scales?
- How can existing practices of image analysis, computer vision, machine learning and visualization be leveraged to address these problems?

5 Meeting schedule

To achieve our goals, the workshop used the presentation format. The purported focus was to be on delineating directions and questions in lieu of free-wheeling discussions and presentations. For instance, all invitees were asked to submit a 1-page response to questions on the formulation of the problem and likely methodologies of integrated omics. These one-page responses were used to organize the workshop with discussions being led by specific senior members. The program reflected the

entire biological scale spanning the genome and its various phenotypic forms and allowed an opportunity to create rubrics for to compare patients as they respond to treatment. Finally, we discussed resources for data and methods. The overwhelming "demand" was to create data assets and then revisit the same questions albeit in a more nuanced manner. Please find the full schedule below.

Sunday, November 11, 2018

15:00 – 19:00 Check-in
19:00 – 21:00 Welcome Banquet

Monday, November 12, 2018

09:00 – 09:30 Welcome & Introductions
09:30 – 10:30 Keynote 1 - Pathology Clinical Questions ([Junya Fukuoka](#))
10:30 – 11:00 Discussion
11:00 – 11:20 Break
11:20 – 11:40 [Andrey Bychkov](#)
11:40 – 12:00 [Kishio Kuroda](#)
12:05 – 13:30 Lunch
13:30 – 14:30 Keynote 2 - Multi-Modality Tissue Imaging ([Rob West](#))
14:30 – 15:00 Discussion
15:00 – 15:20 Break
15:20 – 15:40 [Carolina Wählby](#)
15:40 – 16:00 [Gianni Monaco](#)
16:00 – 16:30 Introduction to open scientific challenges ([Korsuk Sirinukunwattana](#) & [Lee Cooper](#))
18:00 – 19:30 Dinner

Tuesday, November 13, 2018

07:30 – 09:00 Breakfast
09:00 – 10:00 Keynote 2 - Computing Sciences ([Lee Cooper](#))
10:00 – 10:30 Discussion
11:00 – 11:20 Break
11:20 – 11:40 [Jun Xu](#)
11:40 – 12:00 [Korsuk Sirinukunwattana](#)
12:00 – 13:30 Lunch
13:30 – 13:45 Group Photo
14:00 – 14:20 [Guadalupe Canahuate](#)
14:20 – 14:40 [Hirohisa Oda](#)
14:40 - 15:00 [Kazuaki Nakane](#)
15:30 – 16:00 Break
16:00 – 16:20 [Seiichi Uchida](#)

- 16:20 – 16:40 [Saadia Iftikhar](#)
17:00 – 17:50 [Discussion: AI - Expectations & Challenges](#) (Moderator: Junichi Tsujii)
18:00 – 19:30 Dinner
20:00 – 21:00 Software and Tools (open working session)

Wednesday, November 14, 2018

- 07:30 – 09:00 Breakfast
09:00 – 10:00 Keynote 3 - Patient similarity ([Elisabeta Marai](#))
10:00 - 10:30 Discussion
10:30 – 11:00 Break
11:00 – 11:20 [Mitsuyuki Nakao](#)
11:20 – 11:40 [David Wedge](#)
12:00 – 13:30 Lunch
13:30 – 20:45 Excursion and Dinner at “Minemoto”

Thursday, November 15, 2018

- 07:30 – 09:00 Checkout & breakfast
09:00 – 10:00 A Future role of pathology ([Anil Parwani](#))
10:00 – 10:30 Discussion
10:30 – 11:00 Break
11:00 – 12:00 [Summary of challenge ideas](#) (Korsuk Sirinukunwattana & Lee Cooper)
11:30 – 12:00 Closing Discussion
12:00 – 13:30 Lunch

6 Organisers

Dr. Raghu Machiraju led the writing proposal to establish this workshop as part of the NII Shonan Meetings. Due to serious medical emergency Dr. Machiraju was not been able to attend the workshop. With only very short notice Dr. Elisabeta (Liz) Marai agreed to support the organisation of the meeting.

Raghu Machiraju

Institution: Bioinformatics, Computer Science and Engineering and Pathology, The Ohio State University, USA

Contact: raghu.machiraju@gmail.com

Raghu Machiraju is a Professor of Biomedical Informatics, Computer Science and Engineering, and Pathology at The Ohio State University. He is currently serving as the Interim Faculty Lead and Executive Director of Translational Data Analytics Institute. Raghu is also a co-founder of a biotech startup dedicated to the automation of wet the wet laboratory. He has been always been interested in the application of computing to problems in the physical and biological and life sciences. Raghu's current research has led to the development of machine learning methods that include multiple modalities from the clinic and the laboratory (e.g., histology images, clinical and transcript data) and the deployment of machine learning methods to automate protocols wet laboratories. As the Faculty Lead of TDAI, he has engaged in building salient research communities of practice on campus and academic programs in data science and analytics.

Elisabeta (Liz) Marai

Institution: Department of Computer Science, University of Illinois at Chicago, Chicago, USA

Contact: gmarai@uic.edu

Liz Marai is an Associate Professor at the Electronic Visualisation Laboratory at the University of Illinois at Chicago. Her research has been recognised by peers with multiple outstanding research awards, including outstanding paper awards, an NSF CAREER award, and multiple NSF and NIH R01 awards. Two of these NIH R01 awards are for precision medicine projects in head and neck cancer research, and span multiple geographical sites in the United States. She has served as General Chair and Program Chair for the BioVis conference, as an Associate Editor, program chair, and program committee member.

Jens Rittscher

Institution: Institute of Biomedical Engineering & Big Data Institute, University of Oxford, UK

Contact: jens.rittscher@eng.ox.ac.uk

The research of Jens Rittscher is to enable biomedical imaging through the development of new algorithms and novel computational platforms. Current focus of his research is to improve mechanistic understanding of cancer and patient care through quantitative analysis of image data. In 2013 he has been appointed to the first joint academic appointment between the Department of Engineering Science and the Nuffield Department of Medicine at the University of Oxford, UK. He has been awarded the title of Professor of Engineering Science. He is a group leader at the Target Discovery Institute and is an adjunct member of the Ludwig Institute of Cancer Research. Prior to coming to Oxford Jens Rittscher was a senior research scientist and manager at GE Global Research in Niskayuna (NY, USA), one of the world's largest and most diversified industrial research laboratories. Building on his extensive expertise in computer vision, probabilistic modelling and statistical learning, he developed new theoretical approaches that address specific real-world challenge problems in automated video annotation, visual surveillance, and biomedical imaging.

Motonori Ota

Institution: Graduate School of Information Science, Nagoya University, Japan

Contact: mota@i.nagoya-u.ac.jp

Motonori Ota is a scientist in structural-bioinformatics. He started his research on protein folding, structure prediction, protein *de novo* design, and stability change due to mutations. He developed methodologies to analyse trajectories of molecular dynamics simulations, compare protein complexes, detect protein structural change, etc. Also, he developed databases of protein structural changes in the PDB and intrinsically disordered proteins. He is interested in the relationships between protein sequence, structure, function, interactions, and more macroscopic biological phenomena.

7 Talk abstracts

7.1 Junya Fukuoka

Title: Bringing Artificial intelligence to pathology practice. What are the current barriers?

Institution: Department of Pathology, Graduate School of Biomedical Sciences. Nagasaki University

Abstract: Recent rapid progress of digital imaging and AI technologies are enabling a tremendous and significant revolution in medicine. Pathology has served as the diagnostic base of clinical medicine and is best suited for characterizing the underlying driving biological events of most diseases. However, over the last few decades molecular science is increasingly replacing the central role of pathology-based research and diagnosis. Now, given the widespread role and use of digitization and digital slides, pathology-based diagnosis is increasingly used in the clinic. Simultaneously to analyze large digital images from the clinic, AI and machine learning techniques are being increasingly deployed. However, there have been several barriers for a full-scale adoption of digital imaging and AI technologies. First, the digitization of pathology, mostly by whole slide imaging technology, is cost-intensive with a majority of clinical institutions being unable to afford to install sufficient infrastructure. Second, there is a paucity of human resources with sufficient talent. Third, education in the methods of AI require significant efforts on behalf of practicing pathologists, usually overburdened by the daily routine of constantly increasing workloads and a continuing shortage of pathologists. All of these issues should be carefully addressed and mitigate at the earliest possible juncture of an era of integrated pathology and AI.

Short biography:

Dr. Junya Fukuoka is currently working in Department of Pathology, Graduate School of Biomedical Sciences. Nagasaki University, Japan. His research interests include pulmonary conditions, respiratory diseases, and critical care medicine. He serves as a director of the Digital Pathology Association (DPA).

7.2 Andrey Bychkov

Title: Digital pathology workflow in the multi-center environment – Primary diagnosis and beyond

Institution: Department of Pathology, Kameda Medical Center, Kamogawa, Chiba, Japan

Abstract: The Nagasaki-Kameda Digital Pathology (DP) network connecting an academic institution (Nagasaki University Hospital), a large-scale hospital (Kameda Medical Center), and several independent and affiliated centers, was established in 2017. It includes around 40 pathologists, both general and specialized, who are responsible for diagnosing over 40,000 cases annually. As and when the implementation is completed, the network will be effectively employed on the “all day round” basis. Telepathology activities include remote sign-out sessions for primary diagnosis (three per day), tumor boards, multidisciplinary team consultations, journal clubs, research progress meetings, and regular international web conferences. This year we achieved a fully digital workflow for biopsies, surgical specimens, immunostains, outside consults, and frozen. WSI-based education essentially incorporated into all telepathology activities is highly attractive for pathology residents, rotating clinical fellows, and undergraduate medical students. Noteworthy, this DP model considers the immediate adoption of AI/deep learning technologies for diagnostic and research purposes. Currently, we are testing and validating in-house AI solutions for evaluation of tumor content in the biopsies of lung cancer. We will present an overview of the network, which is intended to update the audience about the array of opportunities for computational sciences provided by the DP routine workflow.

Short biography: Andrey Bychkov, M.D., Ph.D. is Director of Digital Pathology at Kameda Medical Center, Kamogawa, Japan. Dr. Bychkov graduated with an M.D. from Russia where he also completed residency training and practiced in anatomic pathology. Later, he earned a Ph.D. in Japan and supervised research projects at Chulalongkorn University, Thailand. He has authored around 50 journal articles and book chapters, serves as an editor for Pathology Outlines and peer reviewer for multiple journals. His research interests are digital pathology and thyroid pathology. Dr. Bychkov is actively engaged in the development and validation of the Nagasaki-Kameda Digital Pathology Network.

7.3 Kishio Kuroda

Title: Digital pathology workflow after pathological AI introduction in clinical use

Institution: Department of Pathology, Kameda Medical Center, Kamogawa, Chiba, Japan

Abstract: The Nagasaki-Kameda Digital Pathology (DP) network is seeking a new pathological diagnosis workflow utilizing whole-slide imaging (WSI) and network connections for the purpose of improving diagnostic accuracy and educating young pathologists. Although pathological diagnosis workflows are expected to change due to advances in WSI and organizational networks, currently it is impossible to effectively utilize WSI-based technologies. It is not sufficient to replace the glass slide by the digital slide. Although AI methods promise tremendous change, they are yet to be adopted in clinical practice. By introducing AI into the clinic not only is the diagnostic accuracy improved, but also tumor enrichment that a typical pathologist is not good at can be standardized. Still, by training semi-supervised deep and machine learning pathologists can contribute to the adoption of AI technologies. Subsequently, adequately trained methods can aid and enhance the diagnosis of pathologists. Mutual education between AI methods and pathologist is necessary for new and effective pathological diagnostic workflows. In this talk, we introduce a pathological diagnosis workflow informed by our AI experience and the problems we faced to improve the accuracy of AI.

Short biography: Kishio Kuroda is an Assistant Professor at the School of Medicine of Nagasaki University. He trained at the University of Toyama and completed this residency programme at the Juntendo University. He specialises in cardiovascular surgery and has a strong interest in applying AI to digital pathology.

7.4 Robert West

Title: Multi-Modality Tissue Imaging

Institution: Stanford University Medical Center, USA

Contact: rbwest@stanford.edu

Abstract: Emerging spatially oriented techniques will provide opportunities to study genomic and proteomic variation within a pathology specimen. This includes generating RNA and DNA sequencing data. Such observations can be correlated with imaging approaches at a single cell or collection of cell level. I will discuss approaches to generating multi-modal spatially oriented datasets.

Short biography: Rob West, MD, PhD, is a Professor of Pathology at Stanford University Medical Center. He is a clinician scientist with experience in translational genomics research to identify new prognostic and therapeutic markers in cancer. His research focus is on the progression of neoplasia to invasive carcinoma. His lab has developed several spatially oriented *in situ* methods to study archival specimens. In addition to running a research laboratory, he also serves as a surgical pathologist specializing in breast pathology at Stanford University Hospital.

7.5 Carolina Wälby

Title: Combining image-based in situ RNA sequencing with quantitative analysis of cell and tissue morphology

Institution: Dept. Information Technology, Uppsala University, Sweden, and Broad Institute of Harvard and MIT

Contact: carolina.wahlby@it.uu.se

Abstract: Typically, tissues are profiled on the RNA level using extracts from tissue homogenates, or in isolated individual cells using single cell capture techniques to resolve gene expression differences among cell types in the tissue. We have developed a targeted sequencing method based on next-generation sequencing chemistry that enables parallel RNA analysis in morphologically preserved cells and tissue. This means that cell and tissue morphology can be quantitatively analyzed and directly related to the local gene expression. We sequence large numbers of individual mRNA molecules in parallel, generating sequence and spatial information at micrometer resolution. Quantitative measurements describing local tissue morphology can be extracted through automated image analysis, including approaches based on deep convolutional neural networks, and displayed as color-coded overlays and outlines on the original tissue data, for example drawing the users attention to tissue regions with disrupted nuclear morphology or irregular gland shapes. We present a free and open-source framework for full resolution image analysis of large images, with the possibility of visual examination and interaction at multiple resolutions. The interface enables seamless zooming and panning, with the option to toggle multiple layers, such as segmentation masks and classification results, on or off. We make use of the strength and flexibility of existing state-of-the-art open-source software for visualization, creating resolution pyramids, image registration and image analysis. This opens up a way of exploring the heterogeneity of tissue data, e.g. differences in gene expression within tumors and in the surrounding stromal compartment.

Short biography: Carolina Wählby is Program Responsible Professor at Division of Visual Information and Interaction, Dept. of Information Technology, Uppsala University, Sweden. Her research group is focusing on the interface between biomedicine, microscopy, and computer science, developing computational approaches for decision support in digital pathology and cytology, personalized medicine and drug screening. She is also Director of the national SciLifeLab BioImage Informatics facility and leads a number of research projects involving large-scale cell and tissue analysis, including deep learning and morphological methods for spatial transcriptomics, funded by the ERC and the Swedish Foundation for Strategic research. She was elected ISAC scholar 2014, received the SBI2 President's innovation award in 2014, and the Thuréus prize in 2015. She is a member of the Royal Society of Sciences at Uppsala and the Royal Swedish Academy of Engineering Sciences. She has a MSc in molecular biotechnology and a PhD in digital image analysis, and carried out postdoc research within genetics and pathology. She was part of the Imaging Platform of the Broad Institute of Harvard and MIT in 2009-2015 and became full professor at Uppsala University in 2014.

7.6 Gianni Monaco

Title: Highly multiplexed imaging of patient tissues with CODEX

Institution: Department of Pathology, University of Basel, CH

Contact: gianni.monaco@unibas.ch

Abstract: Emerging multiplexed imaging approaches offer effective strategies to dissect tissue architecture at a significant resolution. CODEX (for CO-Detection by indexing) is a technology that can image the signal of up to 50 cell markers from a tissue by scanning it iteratively and imaging 3 antibodies at each iteration. CODEX can find applications in almost any biological research context. Immunotherapy is a particularly important field as it has dramatically changed the cancer treatment paradigm with major clinical successes in different tumors. However, the responsiveness to immunotherapy treatments is still not fully extensive as the response is observed only in a subset of cancer patients. Hence, we expect striking insights from using CODEX to study the complex tumor microenvironment of patients treated with immunotherapy. In preparation for CODEX analyses on cancer tissues, we are currently optimizing our research protocols on available patient tissues. Here, we present the workflow of analysis we used and the preliminary results we obtained.

Short biography: Dr. Gianni Monaco recently completed his Ph.D. at the university of Liverpool. During his Ph.D., he spent two years in Singapore working on computational approaches to analyze immunological data in the context of cancer or other ageing diseases. He has experience in the bioinformatics analyses (mainly with R) of data deriving from microarray, RNA-seq and flow cytometry where he applied algorithms for data pre-processing, quality control, clustering, classification, networks modelling and deconvolution. He recently joined the lab of Prof. Alfred Zippelius in Basel, which focus on cancer immunotherapy. Here, he started working on computational approaches in relation to multiplexed imaging.

7.7 Lee Cooper

Title: Structured Crowdsourcing Enables Highly Accurate Convolutional Segmentation of Histology Images

Institution: Department of Biomedical Informatics, Emory University, Atlanta, USA

Contact: lee.cooper@emory.edu

Abstract: While deep-learning algorithms have demonstrated outstanding performance in semantic image segmentation tasks, large annotation datasets are needed to create accurate models. Annotation of histology images is challenging due to the effort and experience required to carefully delineate tissue structures, and difficulties related to sharing and markup of whole-slide images. We recruited 25 participants, ranging in experience from senior pathologists to medical students, to delineate tissue regions in 151 breast cancer slides using the Digital Slide Archive. Inter-participant discordance was systematically evaluated, revealing low discordance for tumor and stroma, and higher discordance for more subjectively-defined or rare tissue classes. Feedback provided by senior participants enabled the generation and curation of 20,000+ annotated tissue regions. Fully-convolutional networks trained using these annotations were highly accurate (Mean AUC= 0.956), and the scale of annotation data provided notable improvements in image classification accuracy.

Short biography: Lee Cooper, PhD, MS, is Assistant Professor in the Department of Biomedical Informatics at Emory University, and Assistant Professor in the Coulter Department of Biomedical Engineering at Georgia Technical Institute and Emory University. Dr. Cooper's research is focused on biomedical data science and the development of computational methods to analyse multifaceted genomic, imaging and clinical datasets. His lab creates scalable machine learning and image analysis software tools to translate cancer data into information that provides scientific insights and that creates translational impact in cancer studies.

7.8 Jun Xu

Title: Computational pathology for precision medicine

Institution: School of Automation, Nanjing University of Information Science Technology, China

Contact: xujung@gmail.com

Abstract: Computational pathology is the integration of digital pathology with advanced artificial intelligence (e.g., machine learning) technology. Its goal is to use a combination of primary sources of data (e.g., pathology, radiology, and laboratory data, etc.) to achieve more accurate disease diagnosis, prognosis, and optimal clinical care. My talk comprises the following three aspects: 1) computational pathology and how computational pathology will contribute to precision medicine; 2) our recent works in developing advanced deep learning based approaches for a) the histopathological image analysis in the cells and tissues level computation; b) radiological image analysis in voxel level computation; 3) Radiomics and PathOmics for a) tumor quantification, b) disease diagnosis and prognosis, and c) predicting the response to therapy.

Short biography: Dr. Jun Xu is currently a full professor at School of Automation, Nanjing University of Information Science Technology, China. He received his M.S. degree in applied mathematics from the University of Electronic Science and Technology of China, Chengdu, China, in 2004, and the Ph.D. degree in control science and engineering from Zhejiang University, Hangzhou, China, in 2007. From 2008 to 2011, he had been a postdoctoral associate and research assistant at Rutgers University in USA. He had been a Visiting Professor at Department of Biomedical Engineering, Case Western Reserve University, USA. He is the deputy director of Medical Image Processing Committee, Jiangsu Association of Artificial Intelligence; committee member of the Medical Imaging and Control of the Chinese Society of Biomedical Engineering and the committee member of the Medical Imaging of the Chinese Society of Image and Graphics. In the past 10 years, he devoted to develop advanced artificial intelligence algorithms for computer-aided detection, diagnosis, prediction and prognosis on cancers. His research interests include computational pathology; digital pathology; medical image computing; deep learning and its application to medical data. He has authored over 30 peer-reviewed journal publications and over 20 conferences papers. He has 10 issued patents and over 15 patents pending in the areas of computational pathology, computer-aided diagnosis, deep learning, medical image computing. His research work has received grant funding from the National Natural Science Foundation of China.

7.9 Korsuk Sirinukunwattana

Title: Predicting consensus molecular subtypes in colorectal cancer using histology images

Institution: Institute of Biomedical Engineering & Big Data Institute, University of Oxford, UK

Contact: korsuk.sirinukunwattana@eng.ox.ac.uk

Abstract: Colorectal cancer (CRC) is a heterogeneous disease with complex molecular subtypes and diverse clinical outcomes. It has become clear that molecular staging adds clinically relevant prognostic and predictive information to the TNM-classification and therefore has a significant impact on personalised risk stratification of CRC patients. Nevertheless, the high costs and complexity of genomic tests are still the major hindrances to bring this technology into clinical routine. Development of cost-effective biomarkers for better stratification of CRC patients is of central clinical importance.

In this work, we developed a deep learning based method to predict the consensus molecular subtype of CRC from whole slide histology images. We applied the algorithm to a cohort of 410 patients diagnosed with advanced primary CRC from the randomised controlled MRC-FOCUS trial. Our naive classifier achieved an AUC of 0.88 (0.86 - 0.9) for predicting the consensus molecular subtype. The results shows the potential of using morphology of primary CRC tumours and their microenvironment as surrogates to molecular assays to predict clinical outcomes and responses to radiotherapy and chemotherapy.

Short biography: Korsuk Sirinukunwattana is a postdoctoral research assistant in Rittscher's group at the Institute of Biomedical Engineering (IBME), Oxford University. His main interest is the applications of image analysis and machine learning in medical research. He has been actively involved in several research projects in computational pathology focusing on colon, breast, and prostate cancers. He completed this doctoral research in the group of Nasir Rajpoot at the University of Warwick and held a postoc position in Andrew Beck's lab at Harvard Medical School.

7.10 Kazuaki Nakane

Title: Classification of degree of differentiation of colorectal neoplasm by the concept of the homology

Institution: Osaka University, Japan

Contact: nakane196508032@yahoo.co.jp

Abstract: Introduction: Recently, a new method based on the homology theory for analyzing histological digital images has been developed. The method evaluates the Betti numbers in a unit area of an image of a colon to determine the region of interest (ROI). The Betti number can be used to assess the degree of connectivity in tissue. Here, we change the binarizing threshold and investigate the relation between the change ratio of the Betti numbers and the different types of cancerous tissue. Materials Method: Colonic specimens were provided by the Osaka Medical Center for Cancer and Cardiovascular Diseases. Data were gathered for internal quality control on a routine basis and all patients gave informed consent for data collection. Results: The calculated results can be approximated by quadratic functions. The distribution of the coefficient on the squared term and the x-coordinates of the vertices are shown. We can see a characteristic distribution for each type of cancerous tissue. Discussion: As the binarizing threshold decreases, the images gradually fade to white and the structure of the tissue is lost. Under the proposed procedure, in areas where the connections in the tissue are tight and clear, the one-dimensional Betti number changes slowly; conversely, where the connections are vague, such as in a background area filled with impurities, it changes very quickly. The state of this change can be considered an expression of the strength of the connectivity and it differs by type of cancerous tissue.

Short biography: Kazuaki Nakane has obtained his PhD at Kanazawa University. He is a guest Associate Professor and a specially appointed Researcher in Osaka University. By using an idea of homology, he has developed a new image analysis method and applied this method to detect cancer lesion. He has also been successful in image analysis of complex images which seem to have mathematical structures.

7.11 Guadalupe Canahuate

Title: Computational pathology for precision medicine

Institution: College of Engineering, University of Iowa, USA

Contact: guadalupe-canahuate@uiowa.edu

Abstract: Identifying similar cohorts of patients is paramount in precision medicine. However, defining a meaningful measure of similarity is a challenging task due to the high-dimensionality, heterogeneity, and sparsity of the data. In this talk, I will present an overview of the methodology we are developing to leverage clustering as a dimensionality reduction approach for radiomic features and oropharyngeal cancer patient data, and the trade-offs of incorporating clustering into survival analysis. We also consider nearest neighbours as a local cluster that can be used to refine global prediction models and build a personalised prediction for the patient.

Short biography: Guadalupe Canahuate is an Associate Professor in Electrical and Computer Engineering at the University of Iowa. Her research focuses on high-dimensional data management and analysis. Her work includes precision medicine approaches for risk modelling of radiotherapy oncologic patients, distributed indexing and query optimisation, compression, and similarity searches. She earned her M.Sc. Degree from The Ohio State University as a Fulbright scholar in 2004 and her Ph.D. in Computer Science and Engineering from The Ohio State University in 2009.

7.12 Hirohisa Oda

Title: Fully convolutional networks for ganglion cell detection on H&E-stained image

Institution: School of Automation, Nanjing University of Information Science Technology, China

Contact: hoda@mori.m.is.nagoya-u.ac.jp

Abstract: Hirschsprung's disease is a congenital disease that lacks ganglion cells in the intestine tissues. Automated detection and segmentation of ganglion cells on HE-stained images of intestine tissues is desired for assistance of rapid operative diagnosis. The fully convolutional network (FCN) is a deep learning technique for region segmentation, and U-net is one of the most prominent FCN structures for medical image processing. However, segmentation of ganglion cells by U-net sometimes not so accurate, especially around blurred boundaries. Our novel network structure, boundary-enhanced segmentation network (BESNet), is focusing on learning textures around boundaries. Experimental results showed that BESNet allowed us to obtain better performance than U-net for ganglion cells.

Short biography: Hirohisa Oda received MEng from Nagoya University in 2014. After working in industry, he started a PhD program of Nagoya University from 2015. His specialisations are image processing and machine learning. His research interests are 1) computer-aided diagnosis using deep learning and 2) cardiac micro CT image processing.

7.13 Seiichi Uchida

Title: Biomedical image analysis as an interesting machine learning target

Institution: Center for Medical Bigdata, National Institute of Informatics, Japan

Contact: uchida@ait.kyushu-u.ac.jp

Abstract: In the research area of image-informatics, significant progress has been achieved by the deployment of recent machine learning technologies, such as convolutional deep neural networks. For biomedical image analysis, however, application of orthodox machine learning strategies is often difficult due to various difficulties, such as ambiguous ground-truth, imbalanced data in the form of true positives and false negatives, small data sizes, etc. I want to discuss the possibility of utilising various machine learning techniques for dealing with those difficulties.

Short biography: Seiichi Uchida received B.E. and M.E. and Dr. Eng. degrees from Kyushu University in 1990, 1992 and 1999, respectively. Currently, he is a professor at Kyushu University. His research interests include pattern recognition, machine learning, and bioimage-informatics. Dr. Uchida is also a guest professor of Research Center for Medical Bigdata, National Institute of Informatics, Japan. He received 2007 IAPR/ICDAR Best Paper Award, 2010, ICFHR Best Paper Award, and many domestic awards.

7.14 Saadia Iftikhar

Title: Where AI meets Drug Discovery

Institution: University of Basel, Switzerland

Contact: saadia.iftikhar@unibas.ch

Abstract: Current advances in digital health and disruptive technologies allow us to gather large patients data, such as clinical data, pathological and omics data, medical and imaging data, health sensors patients data, in the hospital systems every day. These available complex datasets can be used to identify and understand the causes of diseases and then to cure them in a personalised manner/way. The identification of potential targets is the first step in the drug discovery process. In this talk, I will present the work I have been doing on various digital biomarkers identification in different disease models and translating these novel findings using machine learning and computer vision for further clinical applications. I will also discuss the future of artificial intelligence and its impact in the fields of diagnosis and medicine.

Short biography: Saadia is a trained computer vision and machine learning scientist with a PhD degree from Imperial College London. The main focus of her research work always remain on the development of methods and software tools for patients and other biological sample data coming from various conditions and diseases models. She is currently supporting Oncology research cluster of Personalised Health Basel for the Clinical and Research Data.

7.15 Discussion: AI Expectations & Challenges

Moderator: Junichi Tsujii

Institution: Artificial Intelligence Research Center, AIST, Japan

Contact: j-tsuji@aist.go.jp

Short biography: He completed Kyoto University Graduate School in 1973, gaining a PhD in engineering. He first became an associate professor at Kyoto University, then University of Manchester professor in 1988, Tokyo University Graduate School professor in 1995, and principal researcher for Microsoft Research Asia (Beijing) in 2011 before taking on his current position. He serves on the faculty at the University of Manchester.

7.16 Liz Marai

Title: Spatial Patient Similarity in Precision Medicine for Oncological Radiotherapy

Institution: Department of Computer Science, University of Illinois at Chicago, Chicago, USA

Contact: gmarai@uic.edu

Abstract: Precision medicine extends beyond genetic similarity of patients. For example, in most cancer types, the 3D location of a tumour or the spread of disease to the lymphatic system influence not only the type of therapy applied, but also the survival probability and toxicity effects of a patient. Such 3D and 2D factors are therefore important when determining a cohort of similar patients for precision medicine. In this talk I will present recent results from our joint work with head-and-neck clinical oncologists and radiation therapy specialists to develop similarity metrics that account for this type of spatial patient similarity. I will also talk briefly about data challenges in modelling risk prediction in precision cancer therapy.

Short biography: Liz Marai is an Associate Professor at the Electronic Visualization Laboratory at the University of Illinois at Chicago. Her research has been recognized by peers with multiple outstanding research awards, including outstanding paper awards, an NSF CAREER award, and multiple NSF and NIH R01 awards. Two of these NIH R01 awards are for precision medicine projects in head and neck cancer research, and span multiple geographical sites in the United States. She has served as General Chair and Program Chair for the BioVis conference, as an Associate Editor, program chair, and program committee member.

7.17 Mitsuyuki Nakao

Title: Non24 Rhythms of Intelligent Activity Estimated from Tweeting Patterns

Institution: Biomodeling Lab, Graduate School of Information Sciences, Tohoku University, Japan

Contact: nakao@ecei.tohoku.ac.jp

Abstract: Human behaviour is known to be influenced by a circadian clock whose central mechanism is located in the suprachiasmatic nuclei (SCN) of hypothalamic area. SCN has a two-sided structure each of which contains about 10,000 pacemaker cells. The cellular clock mechanism is believed to be underlain by transcription-translation feedback loops involving the clock genes. For humans, rhythms of core body temperature and melatonin are manifest as expressions of SCN activity, whose period is estimated to be a little bit longer than 24hr. However, the light-dark cycle entrains our SCN to exhibit behavioural rhythms such as sleep-wake rhythm with 24hr period. In this study, we investigate tweeting patterns of 1.5 million Japanese tweeters whose tweeting rate is beyond 2 per day on average for 5years. Tweeting pattern subject to the analysis is a series of tweet counts per 10min. There are several clusters of different rhythmic patterns through supervised image classification method. Among them, we found subjects whose tweeting rhythm with a much longer period than 24hr. Since tweeting is regarded as a kind of intelligent activity, our finding suggests that the rhythm of intelligent activity could deviate from the circadian clock whose period is close to 24hr. A possible mechanism is explored by a coupled phase oscillator model.

Short biography: Mitsuyuki Nakao, Dr. Eng., is a professor of Graduate School of Information Sciences, Tohoku University, Sendai, Japan. His current research interests include modelling of biological rhythms from molecular to behavioural level, bio-signal analysis/interpretation, and e-healthcare system.

7.18 David Wedge

Title: The genomic interrogation of prostate cancer

Institution: Big Data Institute, University of Oxford, UK

Contact: david.wedge@bdi.ox.ac.uk

Abstract: Whole Genome Sequencing (WGS) of DNA extracted from prostate cancer tissue can be used to categorise the evolutionary trajectories within sets of cancers. This reveals information relating to the mutational processes that are active within a tumour, prognosis and likely response to treatment. It can also yield insights into the development of heterogeneity with the primary tumour and the mechanisms and processes involved in metastatic spread.

Short biography: David Wedge is a Group Leader in Cancer Genomics at the University of Oxford, Big Data Institute. He previously studied Chemistry at the University of Oxford and Computing and Mathematics at Manchester Metropolitan University. His group use Whole Genome Sequencing to study cancer evolution and heterogeneity. In particular, he has 3 main areas of investigation: heterogeneity within large cohort studies, including the Pan Cancer Analysis of Whole Genomics (PCAWG) and Genomics England projects; processes and mechanisms underlying metastasis across multiple tumour types; the use of novel technologies to study intra-tumour heterogeneity.

7.19 Anil Parwani

Title: Computational pathology for precision medicine

Title: Future role of pathology

Institution: James Cancer Centre, The Ohio State University

Contact: jens.rittscher@eng.ox.ac.uk

Abstract: Digital pathology including whole slide imaging has the potential to make the pathology workflow and reporting a more robust and dynamic process with many advanced features including incorporation of image analysis and computer-aided diagnosis. The process of digitisation should be viewed as an opportunity to truly advance a pathology practice leading to an improvement in the pathology reporting process as well as improving patient care. The objectives of this presentation are to discuss key factors including technology, people, and infrastructure. There will be a discussion on what works today and what the future directions are for this exciting technology.

Short biography: Dr. Anil Parwani is a Professor of Pathology at The Ohio State University. He serves as the Vice Chair and Director of Anatomical Pathology. Dr. Parwani is also the Director of Pathology Informatics and Director of the Digital Pathology Shared Resource at The James Cancer Hospital. His research is focused on diagnostic and prognostic markers in bladder and prostate cancer, and molecular classification of renal cell carcinoma. Dr. Parwani has expertise in the area of Anatomical Pathology Informatics including designing quality assurance tools, bio banking informatics, clinical and research data integration, applications of whole slide imaging, digital imaging, telepathology, image analysis and lab automation. Dr. Parwani has authored over 250 peer-reviewed articles in major scientific journals and several books and book chapters. Dr. Parwani is the Editor-in-chief of Diagnostic Pathology and one of the Editors of the Journal of Pathology Informatics.

8 Challenge Discussion

The concept of a grand challenge was introduced to the group. Grand challenges can accelerate innovation, collaboration, and open up new avenues of scientific exploration. Importantly, they also contribute to generating more reproducible results, where data are shared, results are fairly compared, and methods are openly scrutinised. In recent years we have already seen quite a few challenges that focus on applying image analysis and machine learning methods to whole slide images and address clinically relevant questions. The [Tumour Proliferation Assessment Challenge](#) was launched in 2016 and provides 500 breast cancer cases from The Cancer Genome Atlas. Each case is represented with one whole-slide image and is annotated with a proliferation score based on mitosis counting by pathologists, and a molecular proliferation score. The [CAMELYON17](#) challenge aims to evaluate new and existing algorithms for automated detection and classification of breast cancer metastases in whole-slide images of histological lymph node sections.

The group discussed what specific challenge questions could be posed. The aim would be to stimulate the development of novel methodological approaches while addressing an important clinical question. The group agreed that the challenge question should be open-ended and should not be focused on a well defined pattern recognition task. Multiple types of data should be used and integrated. Still, there was a need to define metrics and outcomes to guide ensuing explorations arising from the use of data.

Several candidate prevailing problems from the clinic were discussed. Predicting the recurrence of low-grade lung cancer could be a suitable topic for a challenge question. Better prediction models would have a clear clinical impact. They could avoid over-treatment in low-risk patients and help manage patients via patient surveillance. The combined institutions of Nagasaki, The Ohio State University, the University of Oxford and Stanford University would have access to a sufficient amount of data to set up such a challenge. It was discussed if a portion of the challenge should focus on non-smokers as they are a particularly challenging group of patients.

During the planning stage epidemiologists, radiologists, oncologists should be consulted to formulate the clinical questions. Katherine Tossas-Milligan (UIC Cancer Center) and David Carbone (James Cancer Center, Columbus Ohio) are candidates with expertise in epidemiology that could be involved.

Dedicated staff would be needed to coordinate a challenge with a number of international partners. Potentially this project would be an ideal opportunity for a senior postdoc who could use the project for a fellowship applications. It is also likely that significant publications will result following the completion of the contest.

9 Workshop Outcomes

The outcomes of the workshop can be categorised into three different groups: research collaboration initiated by participants, the formulation of novel research questions, and activities initiated by the entire group. In the following we will provide selected examples in each of these categories.

9.1 Research collaborations initiated by participants

Below are some representative examples of collaborations that are being pursued. **Cooper & Tsujii.** Stimulated by the AI discussion Lee Cooper has planned a research visit at the Artificial Intelligence Research Center, AIST, Japan. The visit will allow Lee to discuss some of the AI challenges that relate to tissue imaging and analysis with the team.

Rittscher & West. Robert West will broaden his collaboration with Jens Rittscher. It turns out that the common interest go well beyond some of the more immediate technical challenge problems. The probes developed by the West lab can potentially be applied to a number of different projects in Oxford. Shortly after the workshop the group managed to secure funding to start this work.

Monacho & Rittscher. Gianni plans to visit Oxford to learn more about general image analysis methods. Many of the presented tools could be applied to his work on multiplexed tissue analysis.

Iftikkar. Many of the discussions at the workshop seem to related to ongoing projects at the Centre in Basel. Saadia will share some ideas with the overall group and encourage participants to interact with the group in Basel.

Sirinukunwattana & Cooper. Korsuk and Lee will collaborate and coordinate the upcoming challenge. Here they will also discuss how some of the recently developed software infrastructure for working with digital pathology libraries can be effectively used. They will be working with Anil, Jens, Junya and Raghu.

9.2 New research questions

The introduction to clinical pathology which was given by Junya Fukuoka has been instrumental for setting the scene for the overall workshop. Junya motivated a number of questions that relate to very practical clinical questions, most of which relating to lung cancer. Here, it was particularly important to see how pathologists in Nagasaki work in teams across a network of hospitals. Questions and nuances on how such collaborations can be enhanced and accelerated has been discussed. In the context of this concrete setting the use of various software tools has also been analysed. The question on how pathologists can use advanced machine learning in their existing clinical workflows is on of the research questions that should be discussed in one of the future meetings.

The discussion *AI - Expectations & Challenges* which was led by Juninchi Tsujii encouraged all participants to reflect on the ongoing AI discussion. The AI topics that are being discussed in the United States and Japan are in fact different. While most current focus on the application of deep learning is on pattern recognition and classification, the application of deep learning for performing complex simulations is often neglected. Junichi presented some very interesting examples on how they accelerate traditional simulations using more recent machine learning methods.

9.3 Activities initiated by the entire group

The main activity that has been initiated by the entire group is the organisation of a new challenge. Based on the discussion summarised in Section 8 the group has set up a time table for followup. First, a subgroup will work on refining the clinical question. Afterwards, all members will be contacted if they can contribute data to

the defined challenge. In addition, a group will also identify potential ways to fund a junior research fellow to coordinate the challenge.

10 List of participants

First name	Last name	Institution
Andrey	BYCHKOV	Kameda Medical Center
Guadalupe	CANAHUATE	University of Iowa
Lee	COOPER	Emory University
Junya	FUKUOKA	Nagasaki University
Saadia	IFTIKHAR	Univresity of Basel
Kishio	KURODA	Nagasaki University
Georgeta	MARAI	University of Illinois at Chicago
Gianni	MONACO	University Hospital of Basel
Kazuaki	NAKANE	Osaka University
Hirohisa	ODA	Nagoya University
Mitsuyuki	NAKAO	Tohoku University
Motonori	OTA	Nagoya University
Jens	RITTSCHER	University of Oxford
Korsuk	SIRINUKUNWATTANA	University of Oxford
Junichi	TSUJII	AIST
Seiichi	UCHIDA	Kyushu University
Carolina	WAHLBY	Uppsala University
David	WEDGE	University of Oxford
Robert	WEST	Stanford University
Jun	XU	Nanjing University