The Digital Tissue and Cell Atlas and the Virtual Microscope

Jarosław Skokowski[1*], Marika Bolcewicz[1], Kamila Jendernalik[1], Thierry van de Wetering[1], Jacek Gulczyński[1], Anna Lewandowska[1], Leszek Kalinowski[1]

[1] Department of Medical Laboratory Diagnostics, Faculty of Pharmacy, Medical University of Gdańsk (3a Marii Skłodowskiej-Curie Street, Gdańsk, Poland)

* Correspondence author: jaroslaw.skokowski@gumed.edu.pl; ORCID: 0000-0002-3079-3502

Abstract

With the cooperation of the CI TASK (Center of lnformatics Tri-Citry Academic Supercomputer and network) and the Gdańsk University of Technology, the Medical University of Gdańsk undertook the creation of the Digital Tissue and Cell Atlas and the Virtual Microscope for the needs of the Bridge of Data project. In the beginning, an extensive collection of histological and cytological slides was carefully selected and prepared by pathomorphology experts. After processing and digitising, the specimens were sent to servers of the TRYTON Supercomputer, where storing, searching for and scrolling through the images in the Virtual Microscope was made available. The collection consists of twenty thousand high definition images of human tissues and cells accompanied by structurised clinical metadata. Creating a digital atlas and a virtual microscope is an answer to modern education challenges that shape digital competence and are open to modern technologies. The main idea behind the creation of the information tools and digital image data repositories is using them for the purpose of education and as a basis for the creation of new methods of long-distance education. Those resources are shared under the terms of the open Creative Commons license (CC BY-SA), making it possible for teachers, students, and entrepreneurs to use the images safely and process content included in the presented materials without intellectual property infringement.

Keywords: whole slide image (WSI), virtual microscopy, digital atlas of histology, digital microscopy, digital pathology, virtual pathology slides

https://doi.org/10.34808/x55q-sz53_dyr_roz3

Introduction

The modern world is changing rapidly, at a pace never observed before; the time of the pandemic accelerated the pace of those changes, especially in the field of services and telecommunication networks. Restrictions in mobility on the one hand and the process of facilitating telecommunication, on the other hand, caused humanity to swiftly move to the digital reality. On a daily basis, people make use of advanced information technology that they use for remote work or education. Broadband Internet makes it possible to quickly search for and scroll through huge and diverse databases, while cloud solutions allow for scrolling through them on a PC, laptop and a smartphone, even in places where the Internet connection is weak.

The necessity of distance learning, the value of which was verified by the COVID-19 pandemic, is slowly becoming the new paradigm of education. However, a dynamic development of new technologies and the development of science alone shows that knowledge presented to students in some fields has already become outdated when it is being taught. For that reason, it became necessary to support education through allowing for easy and free access to constantly updated and credible digital education resources. Moreover, sharing new technologies and interesting sources of information with the teachers increases their competence and serves as a chance to prepare further generations to function in a digitised world (Koczy, 2020).

The amount and diversity of open data also encourages individuals to search through the Internet on their own and creates the conditions for self-studying. The race in sharing data and encouraging potential users to access created information platforms motivates scientific communities to share their data sources in a way that is attractive and convenient for the user.

For the purposes of the project Bridge of Data – Multidisciplinary Open System Transferring Knowledge – stage II: Open Research Data, the team from Medical University of Gdańsk undertook the creation of the Digital Tissue and Cell Atlas and the Virtual Microscope with the cooperation of the CI TASK IT Centre and the Gdańsk University of Technology (Gdańsk Tech). Carefully selected histological and cytological specimens were gathered by experts in the field of pathomorphology (Greater Poland Cancer Centre, "Copernicus Sp. z o.o.", Medical University of Gdańsk). Specimens were processed, stained and scanned in the MUG's Department of Medical Laboratory Diagnostics (DMLD) and after digitisation, the images were sent to the servers of the TRYTON supercomputer where storing high definition images was made possible. The created repository of Open Research Data (ORD) consists of structurised metadata describing twenty thousand images that can be viewed in the Virtual Microscope. Creating the Digital Tissue and Cell Atlas, along with the Virtual Microscope is the answer to the challenges of modern education that shapes digital competence and is open to modern technologies.

Processing, staining, digitisation

Biological material is delivered to MUG's DMLD as tissue that is not preserved, paraffin tissue blocks or histological specimens. Depending on the type of the received bi-

ological material, particular procedures are implemented in order to obtain high quality digitised images. After being put in histology cassettes, tissue retrieved directly from the patient needs to undergo the process of tissue preservation which, according to the standards, takes anywhere from 24 to 48 hours. The preservation process alone serves to stop the autolysis, which is autodigestion by enzymes present within the cell, and the metabolic processes, thus allowing for the natural structure of the cell to be preserved. A 4% formaldehyde water solution (10% neutral buffered formalin) is a universal preservative that stabilises the structure of the cell by creating methyl bonds (Litwin and Gajda, 2011).

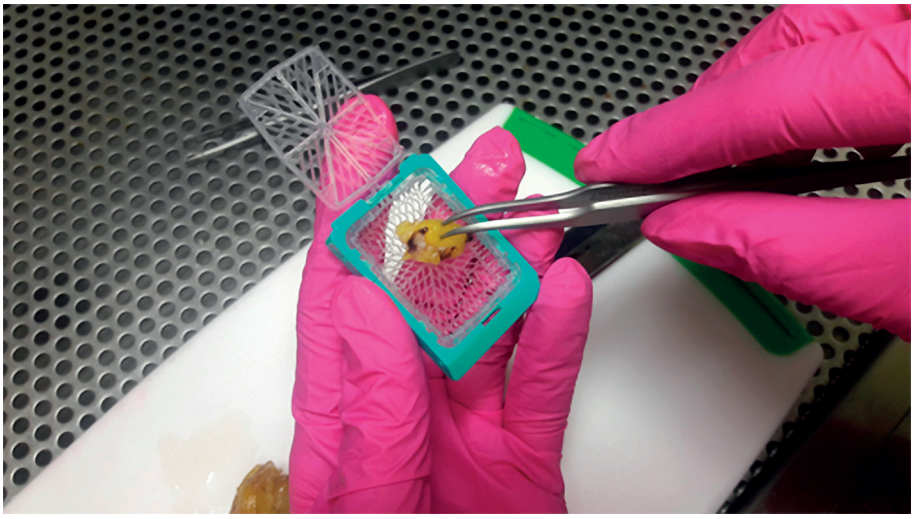


Fig. 3.1. Inserting the tissue into a histological cassette. Kreft, 2020

The tissue prepared in the process described in the previous paragraphs undergoes further steps which are shown in Tab. 3.1. Firstly, the material is processed with the help of an automatised system. The preservative and water are removed from the tissue under vacuum and then the tissue is saturated with liquid paraffin at high temperature. Thanks to modern technology, this process has been reduced from several to merely around 2 hours. Then the tissue is submerged in special cassettes in liquid paraffin which hardens at low temperature. That is how paraffin blocks are created; during this process, the material gains homogeneity and a sufficient degree of hardness which is especially important during further processing (Litwin and Gajda, 2011).

The obtained strips are put in a water bath where the tissue, as well as the paraffin surrounding the tissue, is decompressed and smoothed out, which allows the strip to be fixed on a slide that is later annealed in an incubator set to 60°C. After this step is completed, the slide is ready to be stained by implementing one of the histological staining techniques used in the laboratory (Litwin and Gajda, 2011).

Staining histological slides is a necessary process that makes it possible for the particular cellular structures to be contrasted so that the image achieved under the micro-

scope is legible and unambiguous. In order to achieve that, the slides are stained according to one or a few techniques that will make the specific diagnostic characteristics of the collected tissues visible. The principal technique of positive staining is the hematoxylin and eosin stain (H&E). In this method, the basic hematoxylin dye shows affinity with basophilic structures, such as the nuclei, and stains them blue, while the acidic eosin binds with acidic structures, e.g. cytoplasm, and stains them red. This type of staining provides a general overview of the structure of tissues and cells (Sawicki and Malejczyk, 2012).

After performing H&E staining, additional staining techniques are implemented in order to make specific cellular structures (those in the centre of the attention of a certain researcher) visible. Histochemical staining allows for the identification and localisation of certain chemical substances and biochemical processes taking place in cells and tissues. Substances detected by using this histological technique are for example carbohydrates, lipids, amino acids, proteins and nucleic acids (Sawicki and Malejczyk, 2012).

Methods of molecular biology, such as immunohistochemistry (IHC), immunofluorescence (IF) and a cytogenetic method of fluorescence in situ hybridisation (FISH), are also used in histology (Lewandowska-Ronnegreen, 2018).

The immunohistochemic method makes use of antibodies that bind with antigens present in the tissues. This type of binding is visualised thanks to a coloured detection system directed against the antigen-antibody complex that is created. The result of a positive reaction that confirms the presence of the antigen is brown color visible on the histological slide (Lewandowska-Ronnegreen, 2018).

Just like immunohistochemistry, immunofluorescence also relies on the reaction of the antibody with the antigen. The fundamental difference is a different detection system; antibodies are marked with fluorescent dye and the visual effect of the antigen-antibody complex is glowing in a particular range of the light spectrum (Lewandowska-Ronnegreen, 2018).

In contrast to the IHC method, the fluorescence in situ hybridisation measures the amount of a certain protein; it determines the gene amplification level, namely the number of copies of the gene present in the observed cells. This method relies on the use of gene-complementary probes and the detection system contains probes that fluorochromes are bound to. After a positive reaction takes place, the fluorescent microscope is used to count the signals whose colour depends on the type of fluorochrome used (Lewandowska-Ronnegreen, 2018).

Histological slides stained in the process of the previously presented techniques for the purposes of the Bridge of Data project are later digitised in order to be added to the repository of Open Research Data (Most Wiedzy – The Bridge of Knowledge Research Data Catalogue). During this process, scanners dedicated to this particular goal are used and this creates a digital version of the histological slides. DMDL is in possession of a highly advanced Pannoramic 250 scanner (3DHistech) which combines optical microscope and digital camera functions. The native format of the obtained images is MRXS which, for the purposes of the project, is converted to the universal DICOM format widely used in modern biomedicine while the high resolution of the images is maintained. The Pannoramic 250 camera captures the images of all histological slides with the use of

two available lenses with magnification factors of 20x and 40x and the autofocus function. The device has different types of lighting and diverse observation methods. It allows for scanning microscope positive stain slides in a bright field and additional filters make it possible to capture fluorescent images as well (www.3dhistech.com).

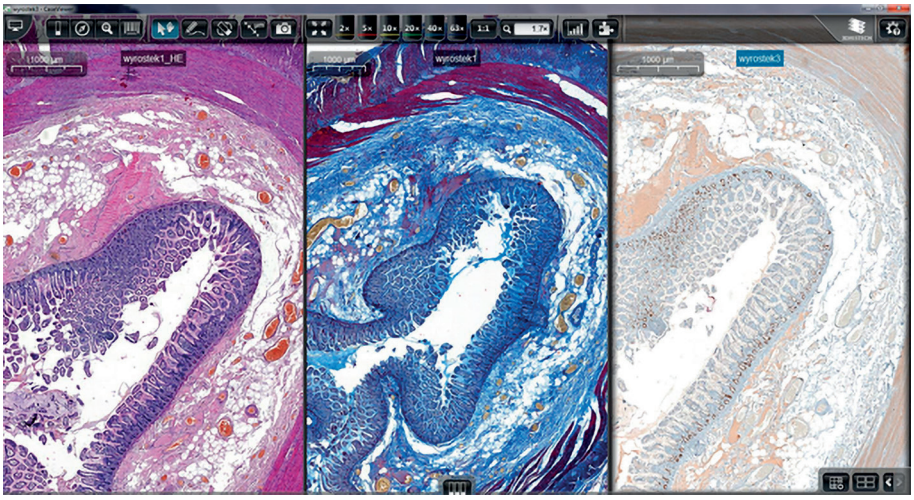


Fig. 3.2. Summary of histology staining techniques. From the left: H&E, HC (Masson's Trichrome), IHC (antigen KI-67). Bolcewicz, 2021

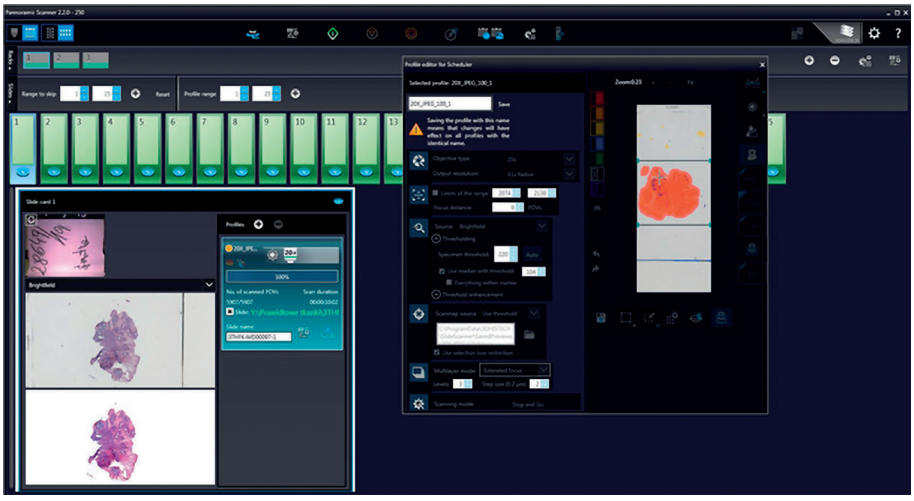


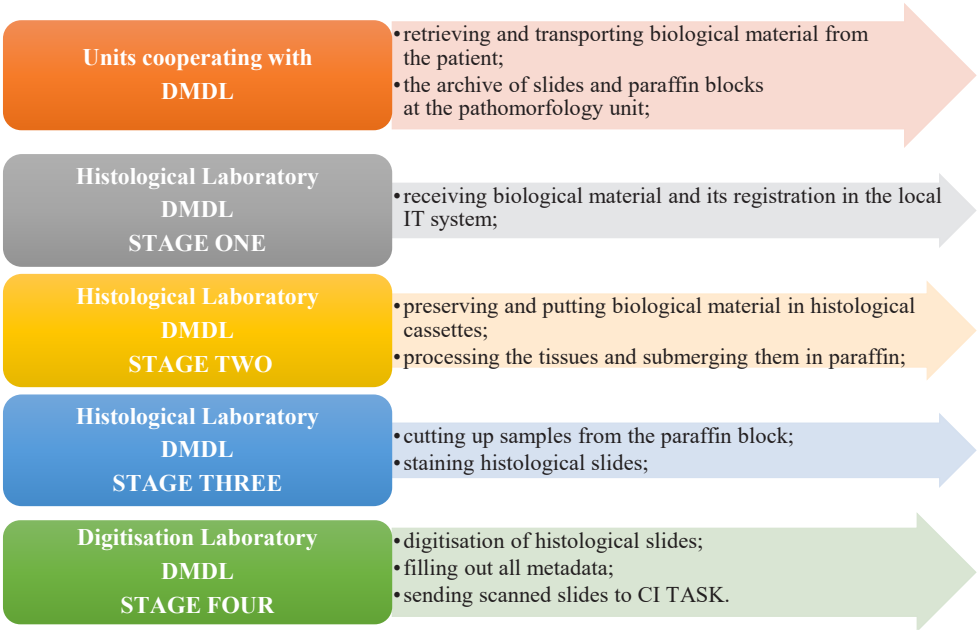
Fig. 3.3. Software supporting the Pannoramic 250 3DHistech histological scanner. Kreft, 2021

Software dedicated to support the scanner is equipped with multifunctional scanning modes that bring out the depth of the tissue, making observation of the image seem like looking at a slide under a light microscope. Additional tools to analyse and process obtained

scans allow for a preliminary assessment to be made according to the standards by the DMDL staff in order to ensure the highest possible quality of images (www.3dhistech.com).

Tab. 3.1

Flowchart of managing biological material in the Bridge of Data project at DMDL



Metadata, sending data to CI TASK

The main goal of the DMDL as a participant of the Bridge of Data project is to create a vast database of image data numbering twenty thousand digitised histological slides along with their detailed descriptions that are jointly called metadata. Metadata constitutes a vital element of every database, because it provides information and describes attributes of particular records. It applies both to technical and research data, allowing for its later identification and further use. However, using metadata requires the preparation of a special key that makes it easier to describe the images. Because of that, the team from the Medical University of Gdańsk created dictionaries with definitions of particular parameters of histological slides and their digital record. They contain attributes assigned to categories that additionally bear an alphanumerical code which is assigned in order to facilitate the procedure of sending data to CI TASK (a project partner).

Metadata used for the project goals was divided into five main categories and contains information such as:

- data concerning the patient (age, sex, the year of a study and a clinical description); information concerning the diagnosis according to the International Classification of Diseases (ICD) and a histological description of the slide, if there are abnormal

images present; histological data concerned with the digitisation process – the type of histological scanner, the lens magnification used and the format of the image; information about the histological slide – the type of the tissue and the material, the way it was collected and the type of staining; additional organisational data – names of cooperating units, the internal number given by the DMDL.

Storing and managing all of the information connected with the digitised resources is possible thanks to a local IT system. After registering the records, a special file with the "yml" extension is generated automatically and contains all metadata essential to describe the digital slides. Packets of data and high definition images are sent with the help of a modern IT infrastructure to CI TASK. A high channel capacity connection (with a speed of 10 Gigabytes per second) allows the images and elements of data to be sent directly to the data centre storage where the created database (after the specialists' approval) will be published in the system of Open Research Data of the Bridge of Data project.

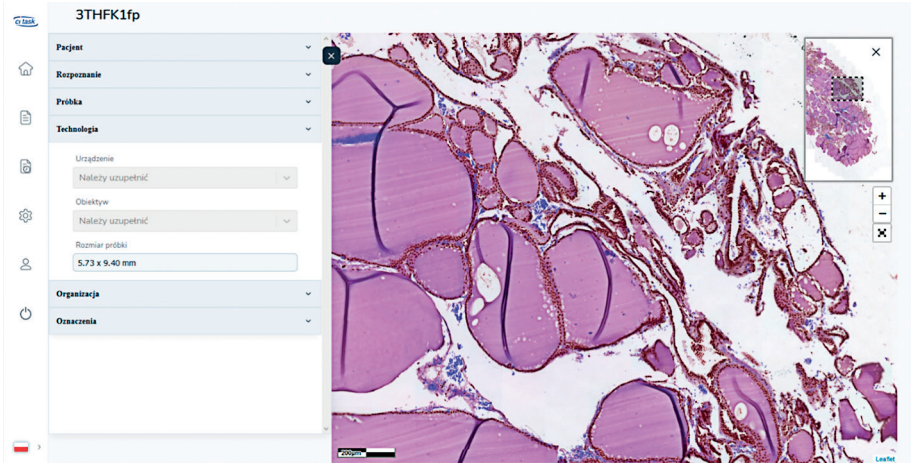


Fig. 3.4. View of thyroid tissue in the Virtual Microscope. Van de Wetering, 2020

Sharing research data – microscopic images

Just like at the beginning of the XIX century, when the development of the optic microscopy and the appearance of histochemical stains complementing the primary H&E stain (hematoxylin and eosin) took place, at the turn of the XXI Century, microscope images started to appear in digitised form. The term "digital pathology" has permanently entered the scientific language for good and the number of publications containing just those two words is steadily rising (around two thousand occurrences in the years 1995–2000 and over eight thousand occurrences during the last 5 years, according to the Pubmed database) (Dee and Meyerholz, 2007). According to the definition, this field of science is concerned with creating a digitised version (digitisation) of a real histological slide and saving it on a physical medium or in a cloud, as well as sharing it in order to analyse or interpret those slides. Whole slide imagining allows for an assessment of the

microscopic image at the place where it was created and consulting any other establishment at the same time which makes it possible to make the final diagnosis in a really short time (Farahani, Parwani and Pantanowitz, 2015). Quick and easy access to those types of images is also extremely important. Additionally they can be connected to LIS (Laboratory Information Systems) where they supply other clinical data of the patients.

The didactic aspect of databases containing digitised microscopic images (both histological and histopathological) is also vital. They can serve as a source of knowledge for the students, medical residents and specialist doctors. Especially those mentioned last can access high definition images of rare cases when paper sources cannot provide images with such resolution parameters. Creating databases in various establishments can also be a starting point for interestablishment research to take place when a pooled analysis of rare cases (occurring only as few or several cases in every establishment) can take place. It helps not only with diagnostics, but also didactics. Of course the important issue of setting threshold parameters arises, if it comes to sharing data, such as the definition of the image, the quality of the scanner and its lenses, as well as the digitised version of the output files, which can constitute a kind of barrier for certain establishments. On the other hand, it enforces actions whose goal is to enhance the quality of work of a digital pathology lab (definition, the number of layers depending on the thickness of the slide, colour scale, output file format, level, degree and image compression algorithm). Similarly not every establishment will be able to purchase a scanner that will make it possible to digitise specimens with non-standard measurements of the microscopic slide. Standardisation and the quality of the microscopic slides alone, as well as routine staining and special staining are also important, but the topic is beyond the scope of this work (Prochorec-Sobieszek, 2016).

Commercial companies concerned with the development of AI (artificial intelligence) will also be allowed to use the created image databases. On this basis, they can develop algorithms for the assessment of digital images which, in the future, will assist pathologists during diagnoses (Prochorec-Sobieszek, 2016).

After evaluating progress in the field of pathology, we can assume that digital pathology will keep making an appearance in other fields more frequently, in order to be used during the process of teaching students and doctors, as well as to assist specialists in the process of diagnosis, and also as a source of valuable scientific material for researchers.

References

Dee, F.R. and Meyerholz, D.K. (2007) 'Teaching medical pathology in the twenty-first century: virtual microscopy applications', *Journal of Veterinary Medical Education*, 34(4), pp. 431–436. DOI: 10.3138/jvme.34.4.431.

Farahani, N., Parwani, A.V. and Pantanowitz, L. (2015) 'Whole slide imaging in pathology: advantages, limitation, and emerging perspectives', *Pathology and Laboratory Medicine International*, 7, pp. 23–33. DOI: 10.2147/PLMI.S59826.

Feldman, M. D. (2015) 'Whole slide imaging in pathology: what is holding us back?', *Pathology and Laboratory Medicine International*, 7, pp. 35–38. DOI: 10.2147/PLMI.S81743

Koczy, S. (2020) '*Nowoczesne technologie w pracy nauczyciela jako szansa na edukację przyszłości*'. Availaible at: https://www.metis.pl/content/view/3456/105/ (Accessed: 8th April 2020)

Lewandowska-Ronnegreen, A. (2018) '*Techniki laboratoryjne w biologii molekularnej*'. Wrocław: MedPharm Polska, pp. 314, 346–351.

Litwin, J.A. and Gajda, M. (2011) '*Podstawy technik mikroskopowych*'. Kraków: Wydawnictwo Uniwersytetu Jagiellońskiego, Wydanie VII, poprawione i rozszerzone, pp. 49–55, 60–68.

Prochorec-Sobieszek, M. (2016) 'Future perspectives of digital pathology', *Journal of Oncology*, 66(4), pp. 277–284. DOI: 10.5603/NJO.2016.0054.

Sawicki, W. and Malejczyk, J. (2012) '*Histologia*', Warszawa: Wydawnictwo Lekarskie PZWL, pp. 4–7.