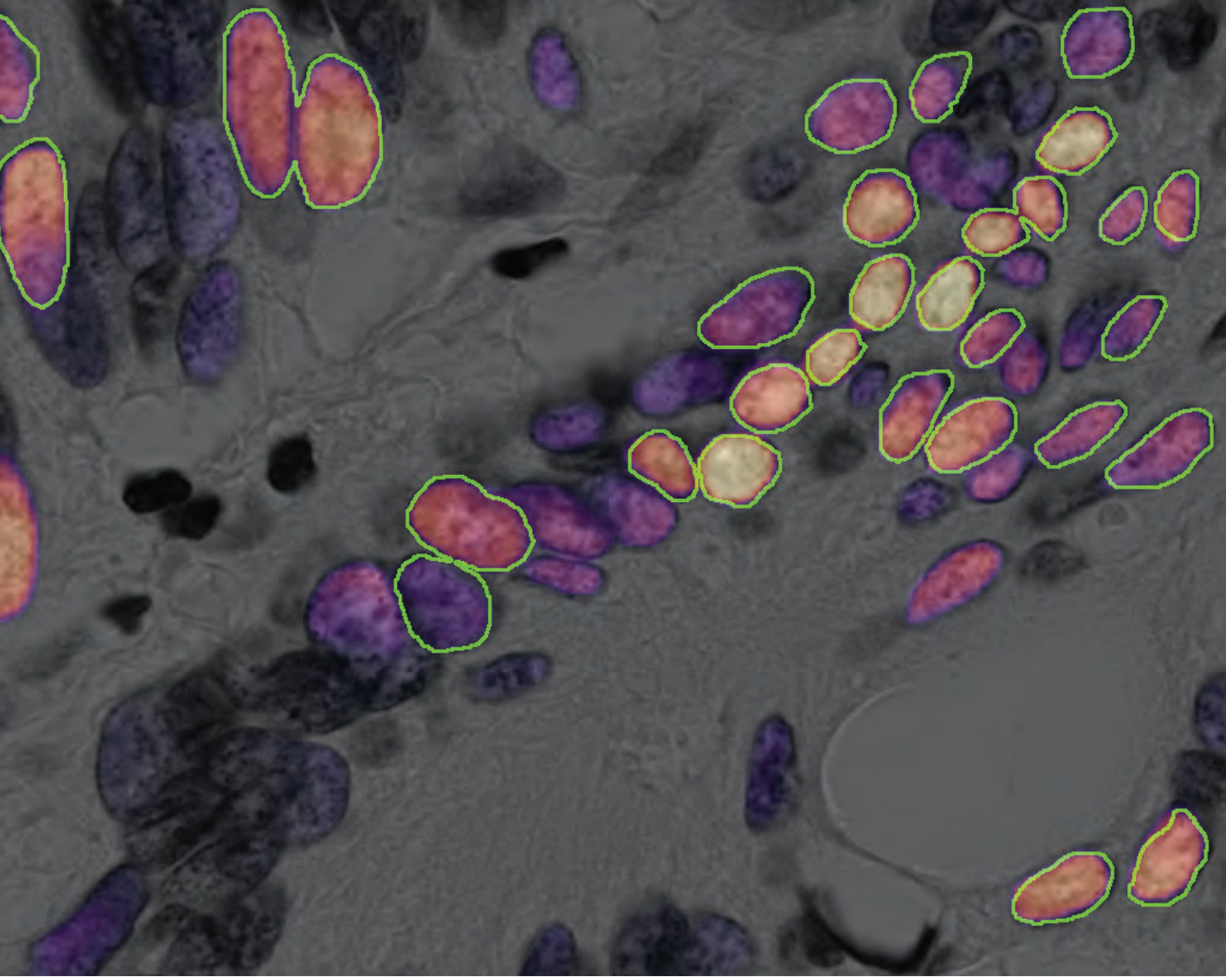


DoMore!

Mid-Term Report

www.domore.no

“

PRECISION MEDICINE

To best treat cancer, we must first understand how the disease will develop. The research in DoMore! will enhance our ability to predict the development of a patient’s cancer, and thereby a more precise treatment for each patient.

The DoMore! project was in 2016 selected as one of the Norwegian Research Council's Lighthouse projects aiming to solve large societal challenges using cutting-edge technology. We have a team of national and international experts within many different fields, including digital image analysis, tumor pathology, cancer surgery, and oncology.

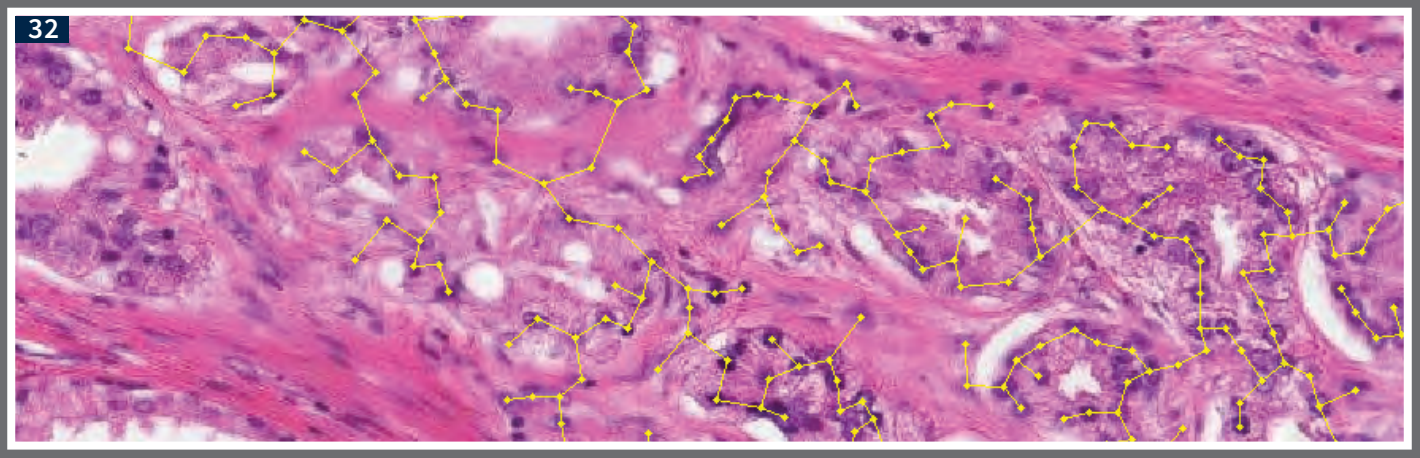
An essential element of the Lighthouse projects is to work in cooperation with users, public and commercial players to convert knowledge into clinical applications. A basis of our selection was the potential for value creation, products, and spinoffs that could produce the greatest impact. The DoMore! project will run until the summer of 2021 under the Director of the Institute for Cancer Genetics and Informatics (ICGI) and DoMore! Project Manager, Professor Håvard E. Greger Danielsen.

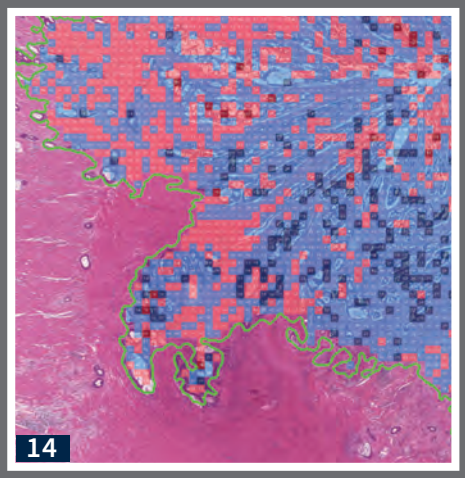
FROM HUMAN DECISION-THINKING TO ARTIFICIAL INTELLIGENCE

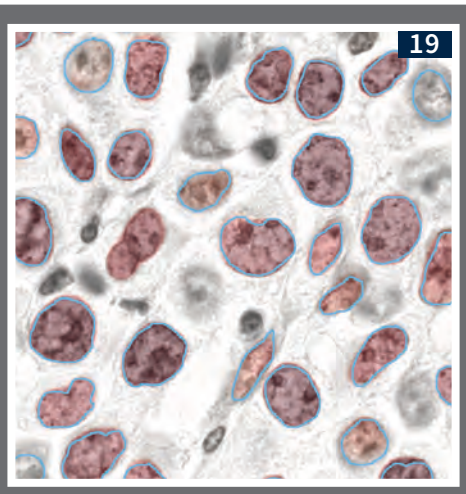
We have since the beginning of the project been working to utilize new technology to improve prognostication of cancers using digital tools for pathology. We have been developing a complete transferal of complex human decision-making from its current basis in visual observation to a computer basis by the use and development of methods based on artificial intelligence (AI).

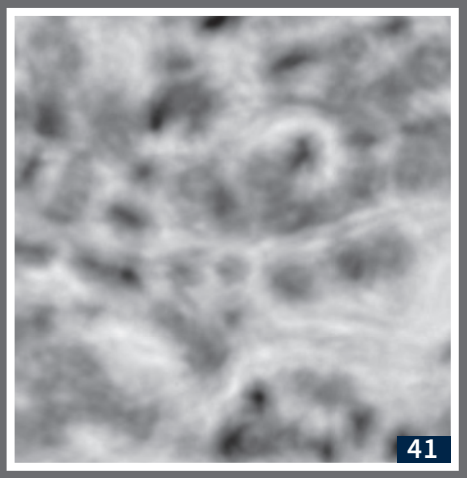
We are basing our concepts on image analysis and more specifically on deep learning, texture analysis, and DNA analysis. We are studying the effect of sampling in the prostate-, lung- and colorectal cancers, and on histological grading, DNA measurements, and gene expression analysis. The experiments described in this report are designed such that multiple samples from each patient and tumor are examined with the goal of modeling the heterogeneity in prostate-, colorectal- and lung cancer and analyzing the effect of sampling on the strength of the prognostic markers. At the end of 2018, we reached our mid-way point of the 5-year long project. The project description has been updated to reflect the current partnerships and project plans, which are detailed in the pages that follow. The general results obtained so far will be reported here, but for specific results that are, or are to be, published in scientific journals, we have to refer the reader to these publications.

The goal has been to increase the number of diagnostic and prognostic tests for cancer patients to provide a more accurate prognosis for the patient. Our results so far show that it is, in fact, possible to teach a computer, not only to do the same but through Deep Learning and Big Data, to establish more robust grading systems in cancer types where pathology is less successful, while at the same time eliminating the subjective component.

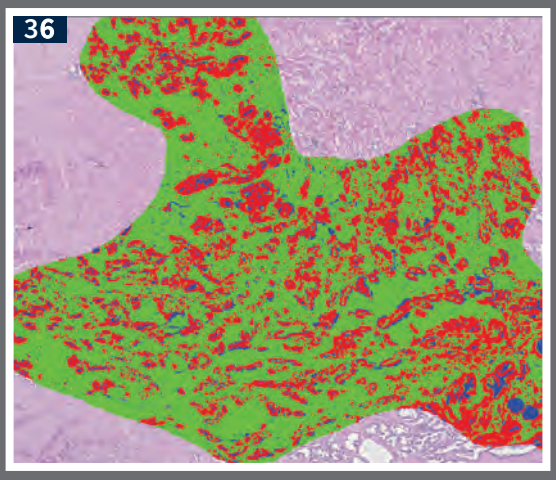

32


14


19


41


Nikolai
Like This
I dag er det verd
alle. Noen opple
familie eller venr
pårørende eller
av ti nordmenn k
behandling og g
jeg besøkt Oslo
intelligens for å
treffsikker kreftb
handler om: Å gi
467
Like
Most relevant
Ragnhild
52

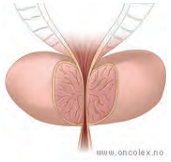

36

CONTENTS

WP 1

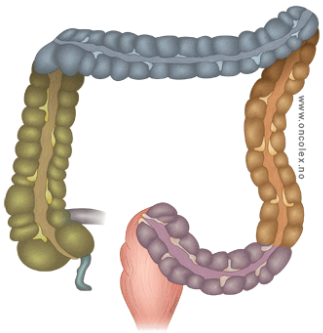
DATA PRODUC-TION

The main goal for DoMore! is to develop generic and objective digital prognostic markers for cancers in the prostate, colon, rectum and lung.



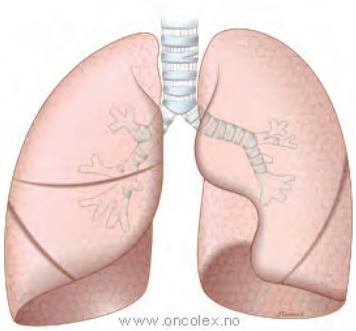
PROSTATE

1 259
4 687
16 437



COLORECTAL

4 099
8 812
24 140



LUNG

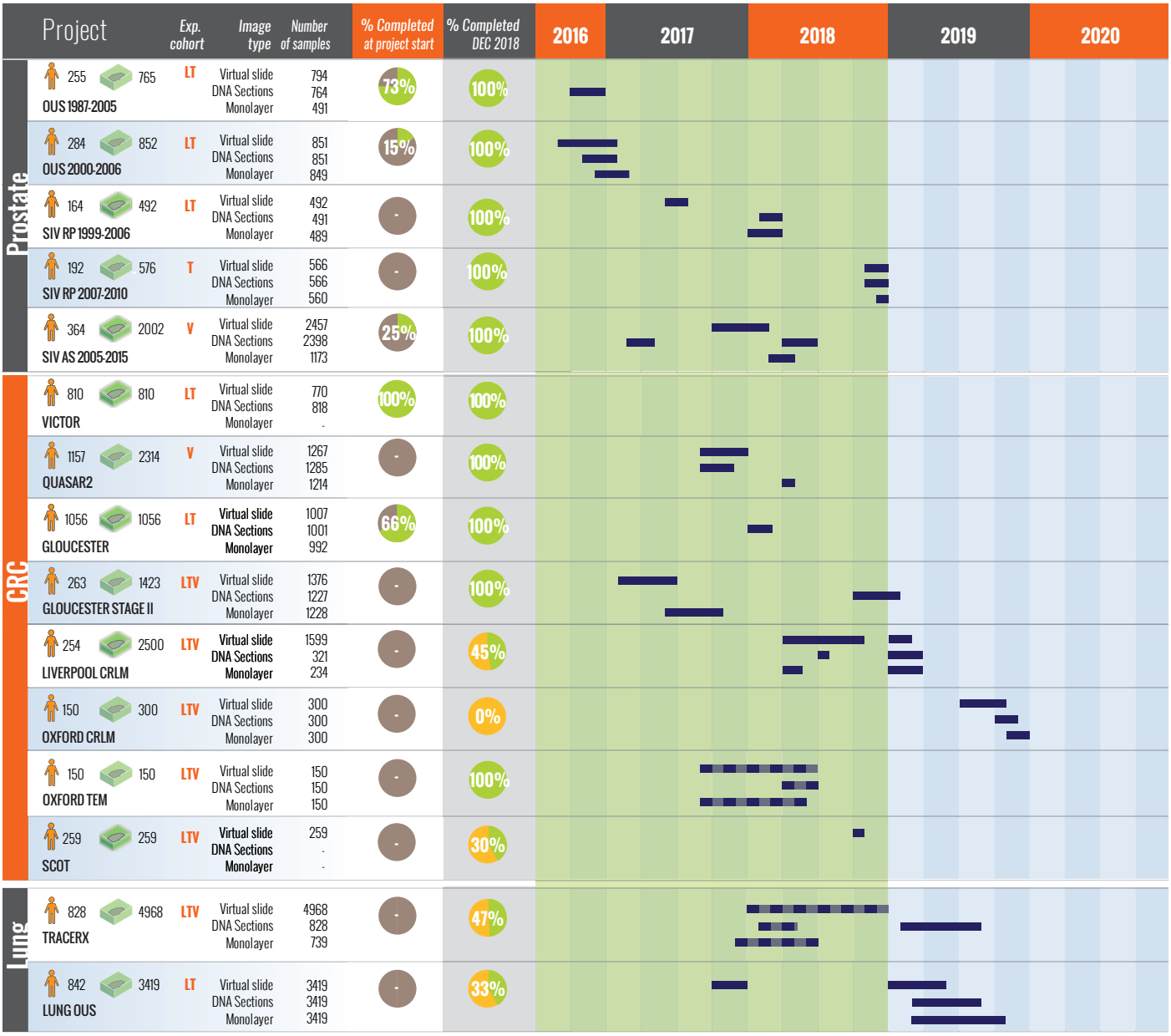
1 670
8 387
25 791

Patients Tissue Blocks Samples

Tumors are heterogeneous with regions containing different genetic or epigenetic aberrations, which pose a great challenge for the prognostication of cancer. The biomarkers used to identify the aggressiveness of a tumor are not distributed evenly throughout the tumor, and by sampling only a minuscule amount of the tumor, there is a great risk of missing the cells that might lead to a patient's death.

The experiments in this project have been designed such that multiple samples from each patient and tumor are examined with the goal of modelling the heterogeneity in prostate, colorectal, and lung cancer and analyzing the effect of sampling on the strength of the prognostic markers.

Work package 1 (WP1) has been responsible for the production of all samples to be analyzed in DoMore! (Progress is shown in the Gantt chart below). The DoMore! project so far includes more than 66 000 samples of 2 886 tissue blocks from 7 028 patients.



Deliverables for WP1 were as follows: For each tissue block or biopsy included in the DoMore! project the lab was to produce 1-4 virtual slides, one DNA section and one monolayer.

VIRTUAL SLIDE

A virtual slide is a scanned H&E stained 3µm tissue section, whereas the virtual slide production line in the laboratory also involves all tissue sectioning for all image types.

Virtual slides are used in WP2 for Histotyping (tumor grading), WP3 for tumor delineation and in WP4 for quantification of stroma, Gleason grading, mitotic index and microtracking (co-analysis of different features on a cell-to-cell basis).

Each block is sectioned, obtaining (at least, but not limited to): one 3µm section, two 5µm thick sections, and one 50µm section. The 3µm section is stained with Hematoxylin and Eosin (H&E), and then our in-house pathologist marks the tumor area in each section. Using the H&E as guidance, the same area on the paraffin block is marked, with a sharp tool. When sectioning the 50µm scroll, we only collect the area marked as the tumor area. This is followed by a new 3µm section (stained with H&E), which the pathologist uses as a control section to verify that there is still tumour tissue left.

In order to develop methods that are invariant o imaging devices, all slides are scanned using scanners from at least two ajor vendors (Leica, Hamamatsu). Also, for the tumor delineation project in WP3, we needed tissue sections both with the pathologists marking of tumor area, as well as the same tissue section without any tumor marking. Based on this, we have so far produced 59 893 individual scanned H&E sections (Virtual slides) in DoMore!

DNA SECTION

DNA sections are used in WP3 for automatic segmentation of cell nuclei, and in WP4 for transation of DNA ploidy and Nucleotyping analysis from monolayers to tissue sections.

5µm sections are first stained with H&E and scanned, then de-stained and re-stained with Feulgen, a DNA specific stain, before it is scanned again. Our in-house pathologist marks the tumor area on the H&E section, the H&E and Feulgen stained sections are then overlaid using the marked tumor area on the H&E section to guide the measurement of DNA in the Feulgen stained section. The final measurements are initially performed on an automatic microscope system developed at the institute. We are currently also exploring the possibilities of replacing these microscope systems by scanners.

MONOLAYER

Monolayers, slides with isolated cell nuclei, are used for DNA ploidy and Nucleotyping analysis (see WP4).

From a 50µm thick tissue scroll cell nuclei are isolated and spun onto a glass slide. The nuclei are then stained with the DNA-specific Feulgen stain, and measured on an automatic microscope system developed at the institute.

THE MATERIALS

All the prostate cancer patient material in DoMore! are from hospitals in Norway; The Oslo University Hospital (OUS 1987-2005, OUS 2000-2006) and Vestfold Hospital Trust (Sykehuset i Vestfold, SIV) (SIV RP 1999-2006, SIV RP 2007-2010, SiV AS 2005-2015). These patient samples have been readily available to us, and the lab has completed all the deliverables related to these patient materials. We lack some recent samples from patients enrolled in the Active Surveillance program at Vestfold Hospital Trust, but are continuously including these as the patients are getting biopsied as part of their follow up.

For colorectal cancer, all the patient materials included in DoMore! are provided by our collaborators in England; The University of Oxford (VICTOR, QUASAR2), Cheltenham General Hospital (Gloucester, Gloucester Stage II), Royal Liverpool University Hospital (Liverpool CRLM) and John Radcliff Hospital (Oxford TEM). Acquiring tissue from collaborators abroad has provided some challenges. Due to, among others, the newly implemented GDPR regulations, there have been delays in getting all legal aspects of transferring tissue to and from Norway in order. However, by being very flexible and quickly reprioritising tasks when needed, the laboratories have throughout the project maintained a continuous workflow.

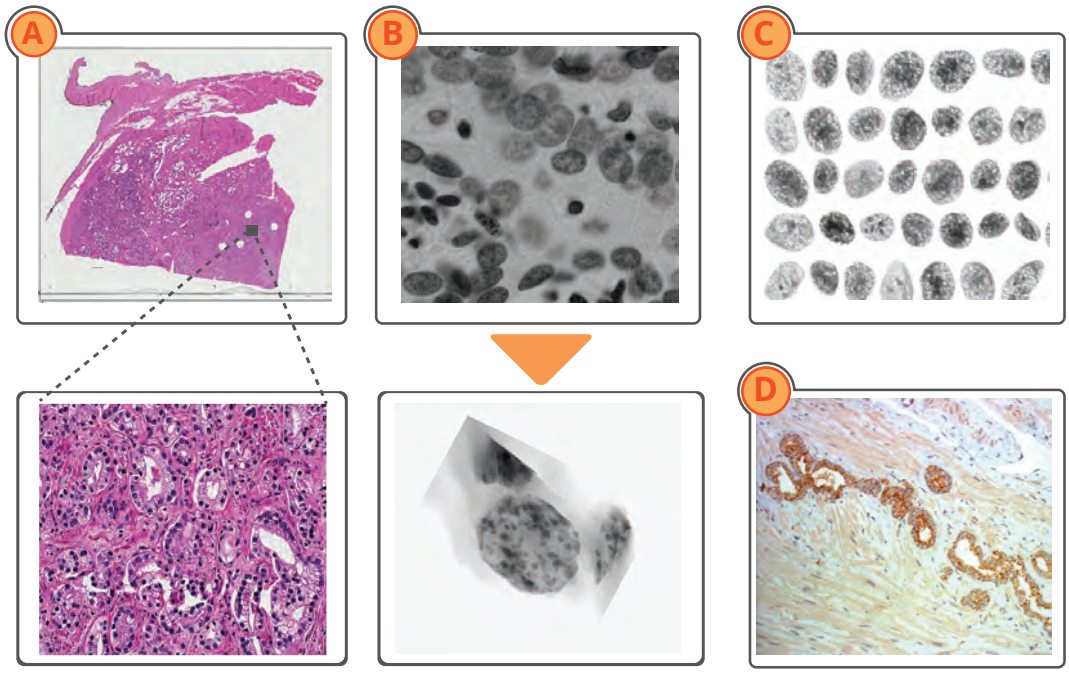
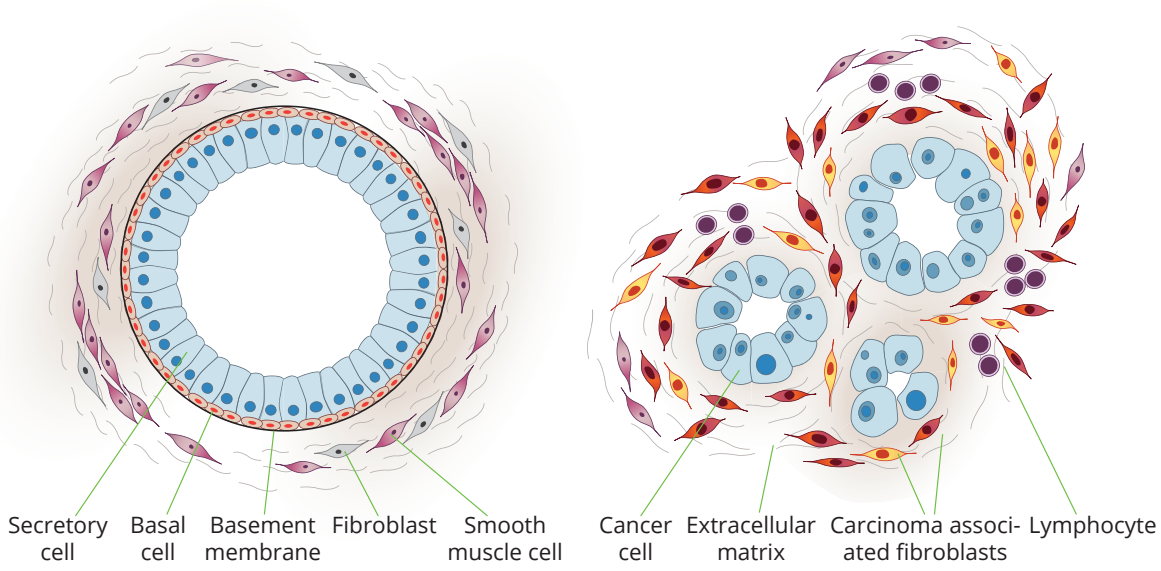


Image types Virtual slide (A), DNA section (B), Monolayer (C) and gene expression shown with antibody binding to protein (D).

One of the prostate cancer patient materials (Nijmegen/Oxford) and two of the colorectal patient materials (AKER1, AKER2), initially included in DoMore! turned out to be difficult to obtain. As a substitute for this, we included three new colorectal patient series (Gloucester Stage II, Liverpool CRLM, Oxford TEM). We are still seaking the replacement for the missing prostate material.

We have completed the deliverables for all the colorectal patient series, except for the patients in Liverpool CRLM. This patient material was included in 2018, and we are also expanding this material to include both primary tumors and liver metastasis from these patients. We expect to finish all deliverables from the Liverpool CRLM by Q2 2019.

The lung cancer patient material in DoMore! are from the Oslo University Hospital (Lung OUS) and University College London (TRACERx). During the project period, it became clear that UCL would not be able to provide all tissue samples as initially agreed upon. The DoMore! project will still receive tissue samples from the TRACERx trial. However, there will be fewer samples available than we initially planned for. To maintain the same amount of included patients and tissue samples we have obtained a new lung cancer patient series from Oslo University Hospital (Lung OUS).



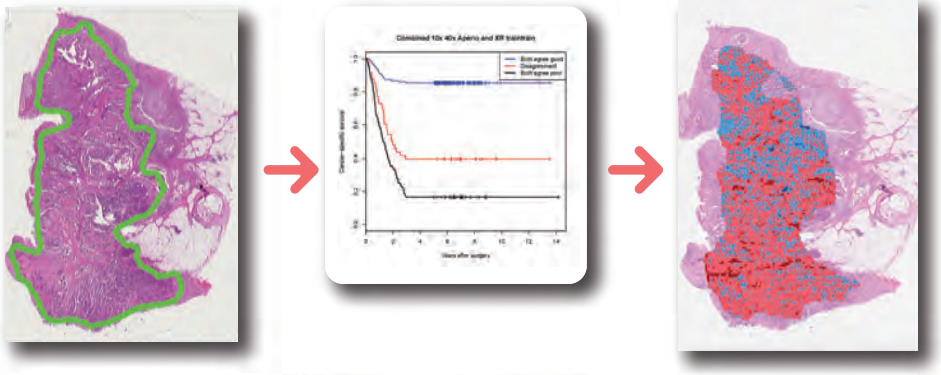
Left: Normal Prostate gland development. Right: Prostate cancer development.

WP2

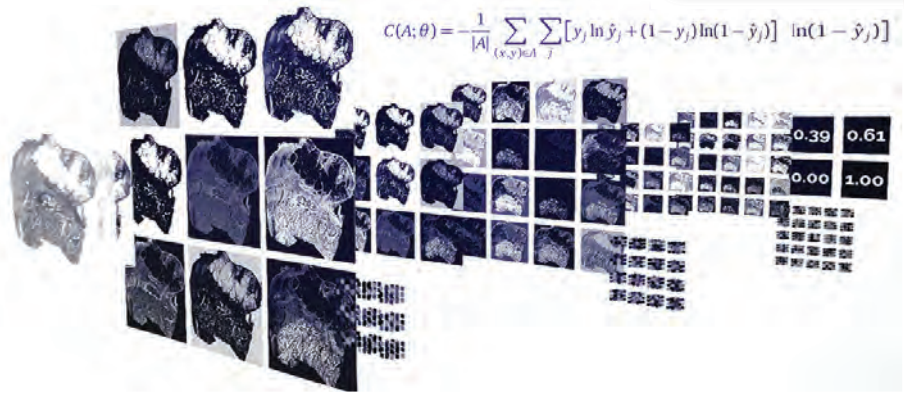
HISTO-TYPING

Histopathology refers to the microscopic examination of tissue to study the manifestation of disease. In the DoMore! project, all patients have been diagnosed with cancer, and their tumours were surgically resected.

Histotyping



Deep convolutional neural networks



The histopathologists study the appearance of the resected specimen and assess relevant properties, such as the tumour grade. Tumour grading is an evaluation of the extent to which tumour cells and tumour tissue resemble normal cells and tissue, where a high degree of similarity (well differentiated) is associated with a better prognosis for the patient than a low degree of similarity (poorly differentiated). The analysis is carried out in HE-stained tissue sections. Tumour grade is a good prognostic marker, but a substantial proportion of patients are classified as moderately differentiated, i.e., an intermediate group with an intermediate outcome. The increasing workload for pathologists as well as significant intra- and interobserver variability implies a need for automated methods for this task.

AUTOMATION

We have developed Histotyping, a fully automated histological characterisation of HE-stained sections from cancer specimens for prognostic purposes. The method is based on deep learning by convolutional neural networks trained on images of HE-stained tissue sections where the patient outcome is used to guide the training process into a system that is able to identify tissue patterns in the HE-sections that are distinct for patient prognosis. The resulting computer model can be applied to a new patient's tumour sample and estimates the probability of a poor patient outcome.

RISKS AND PREVENTIONS

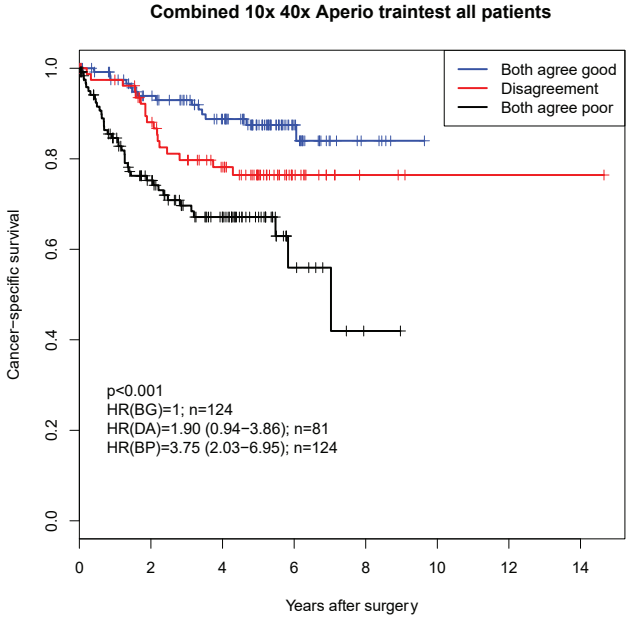
Large amounts of labeled data are required to train this type of deep convolutional neural network and not until recent years have we had the required datasets and computational resources to perform these types of analyses. The neural network models have millions of features and overfitting to the training dataset is a common problem, i.e., that the computer model identifies and exploits artifacts in the training dataset that are associated with the desired outcome, but have no biological relevance and fails when evaluated on a new dataset on which it is not trained. To increase the probability that the neural network generalises when applied to new patients from a new dataset, we have used a robust design with thousands of patients from different patient cohorts with the same cancer type (Table 1). We have developed the framework on stage I-III colorectal cancer patients from two hospitals in Norway and from two clinical trials in England. Another risk in the development of such models is the adaption to the technical equipment such as the imaging system, i.e., that the method works well on images scanned with the scanner on which it is developed and not on images from a scanner from another vendor. To compensate for this problem, we have scanned all images with scanners from two major scanner manufacturers (Hamamatsu and Leica).

Patient cohort	Number of patients	Purpose
Ahus	160	Develop the method
Aker	578	Develop the method
Gloucester	979	Develop the method
The Victor trial	828	Develop the method
Samples from Gloucester prepared externally	979	Test the robustness to technical variation
SCOT	250	Test the method
Colorectal samples from the Cancer Genome Atlas (TCGA)	616	Test the method in a technically very different dataset
Quasar 2	1140	Validate the method

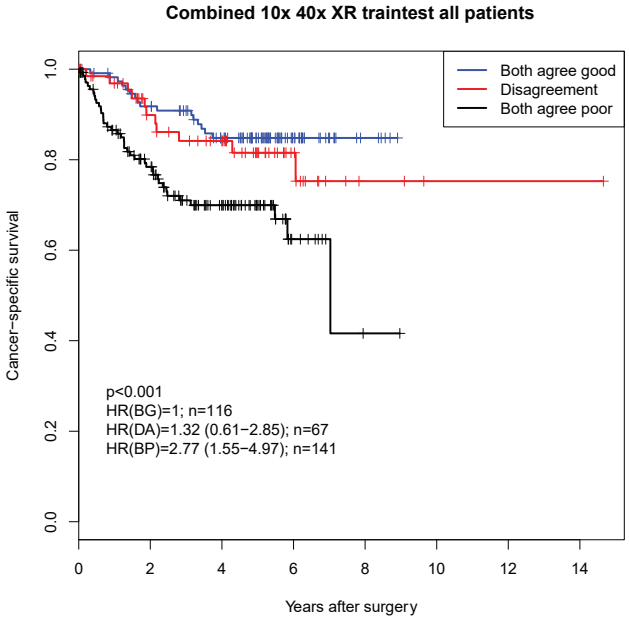
Table 1: Colorectal cancer patient cohorts and purpose in the development of Histotyping

A third risk in the development of a computerised system for risk assessment based on scanned HE-sections is the dependence on lab preparation, i.e., that the method works on scans of tissue sections prepared and stained in the lab where the method is developed only. To evaluate this, we have scanned parallel sections from one of the patient cohorts (Gloucester) that have been prepared in the pathology routine there.

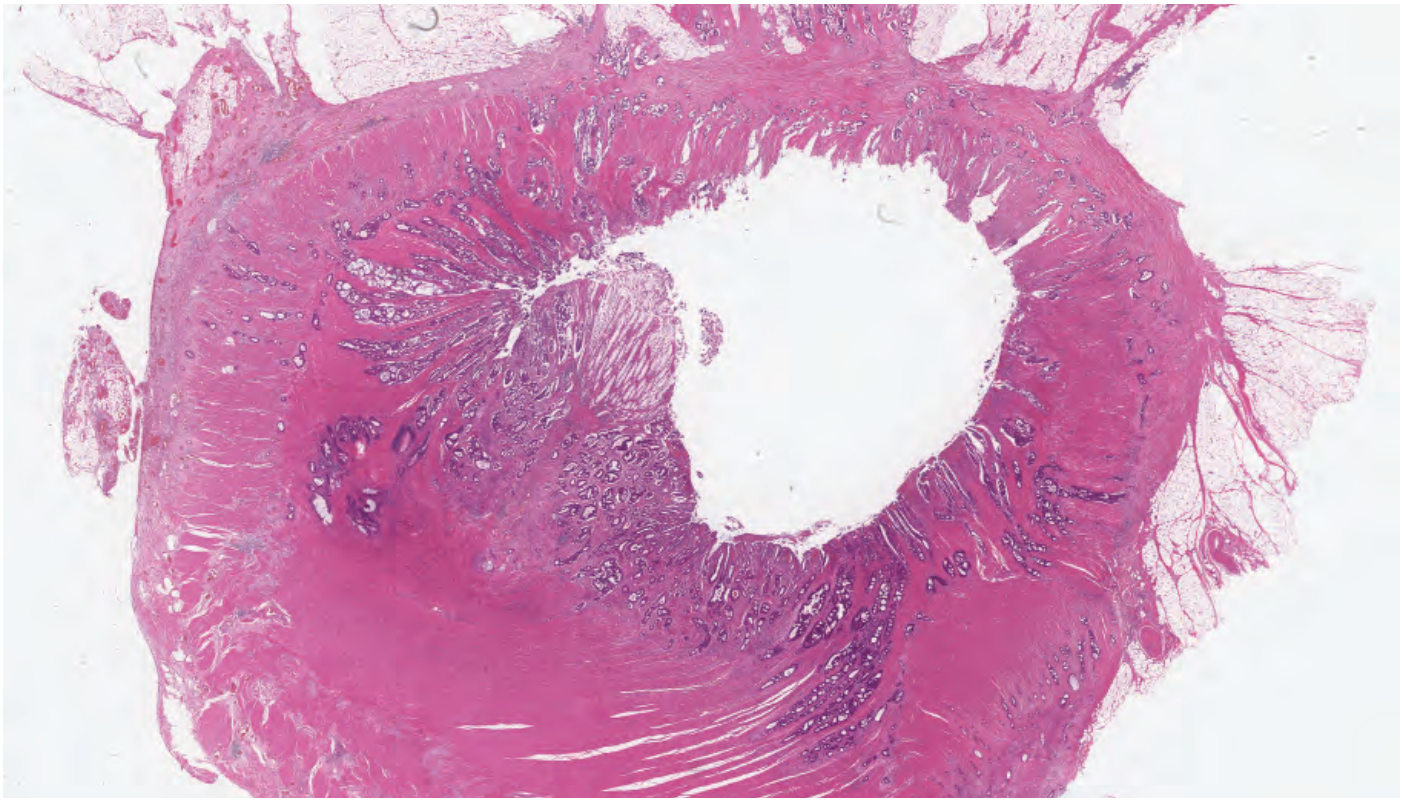
We have partitioned our patient datasets to make the best use of the data. The Kaplan-Meier plots below illustrate results based on images scanned with 10x, and 40x lens analyzed in corresponding neural network models and classified according to the agreement between the two models' classifications in a test partition that has not been included in the training process.

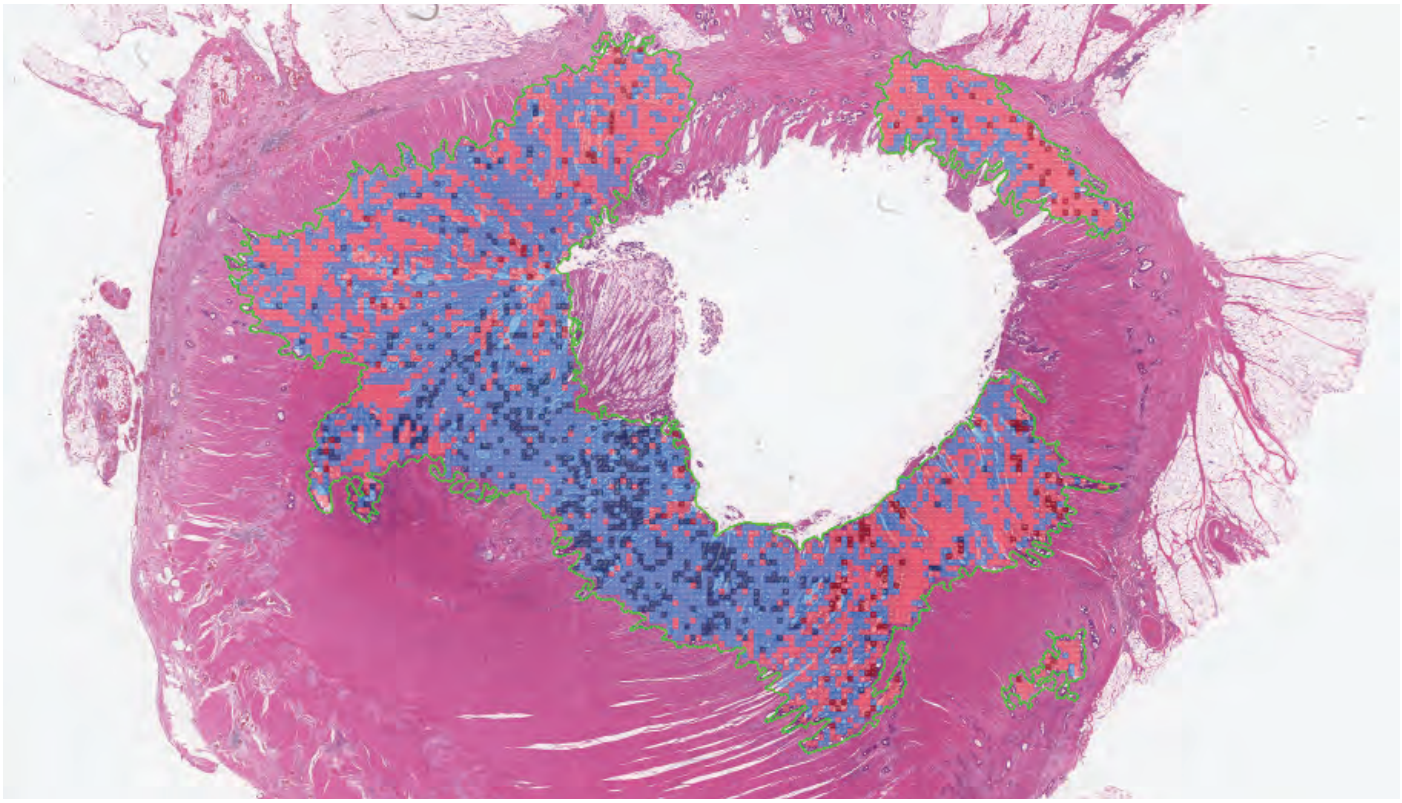


Kaplan-Meier plot of cancer-specific survival probability grouped by classification by 10x and 40x models and their agreement based on scans from the Leica Aperio scanner.



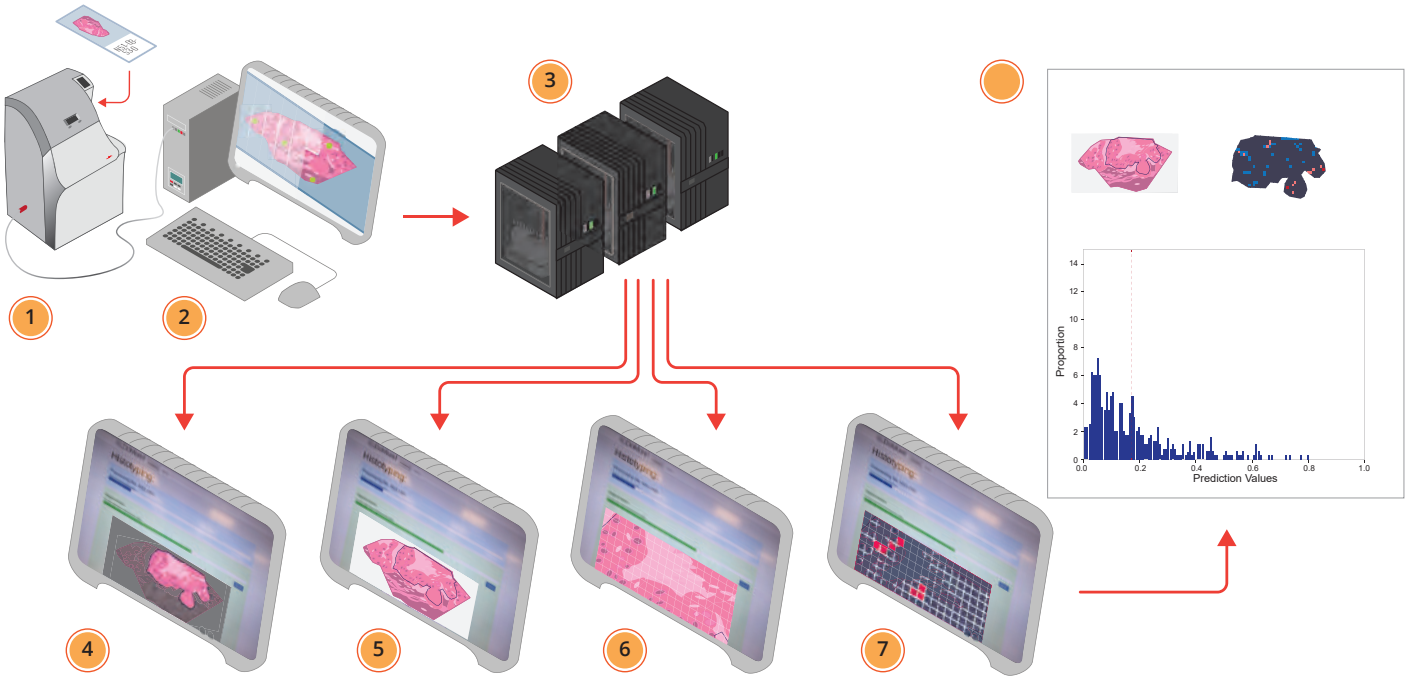
Kaplan-Meier plot of cancer-specific survival probability grouped by classification by 10x and 40x models and their agreement based on scans from the Hamamatsu XR scanner.





HISTOTYPE

Any of a range of tissue types that arise during the growth of a tumour



Histotyping: 1) Samples are fed into the scanner. 2) The scanner scans the tissue sections, adjusting focus to get the best result possible. 3) The scans are sent to the deep learning models, running on GPUs where 4) the tumour region in each scanned sample is automatically segmented, 5) the outlined tumour is further, 6) divided into tiles, and 7) automatically assigned a probability for representing poor prognosis. Each tile is toned blue for good, or red for a bad prognosis. 8) Finally a report is generated, where patient prognosis is estimated based on the individual tiles' prediction values. The different tumour regions' contribution to the patient classification are illustrated as demonstrated in the image on the left page.

WP3

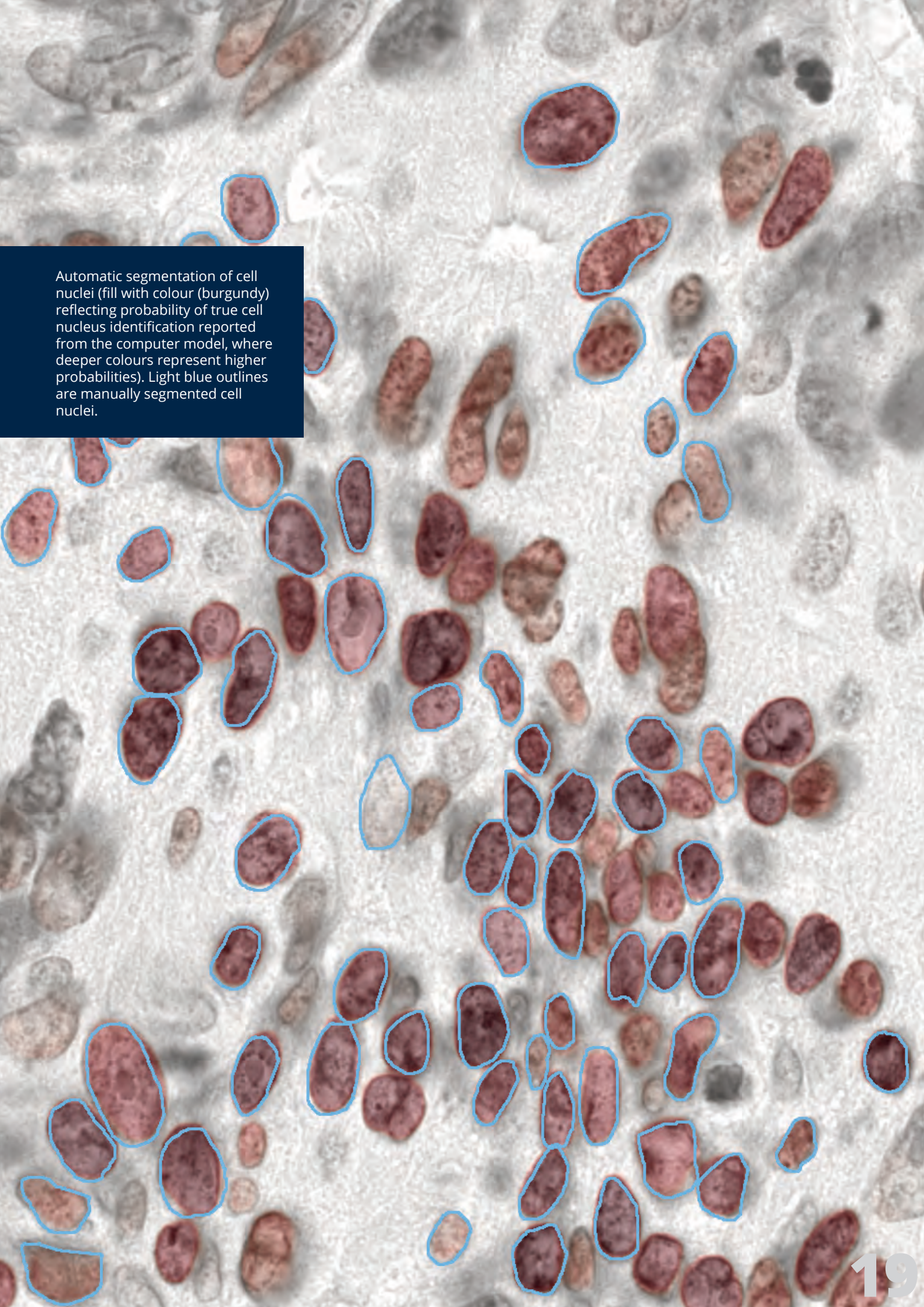
SEGMEN-TATION

The automatic identification of cell nuclei in Feulgen-stained histological sections and automated detection of tumour regions in HE-sections forms the basis for a majority of the DoMore! applications. These two tasks are different and treated in separate projects, described in the following.

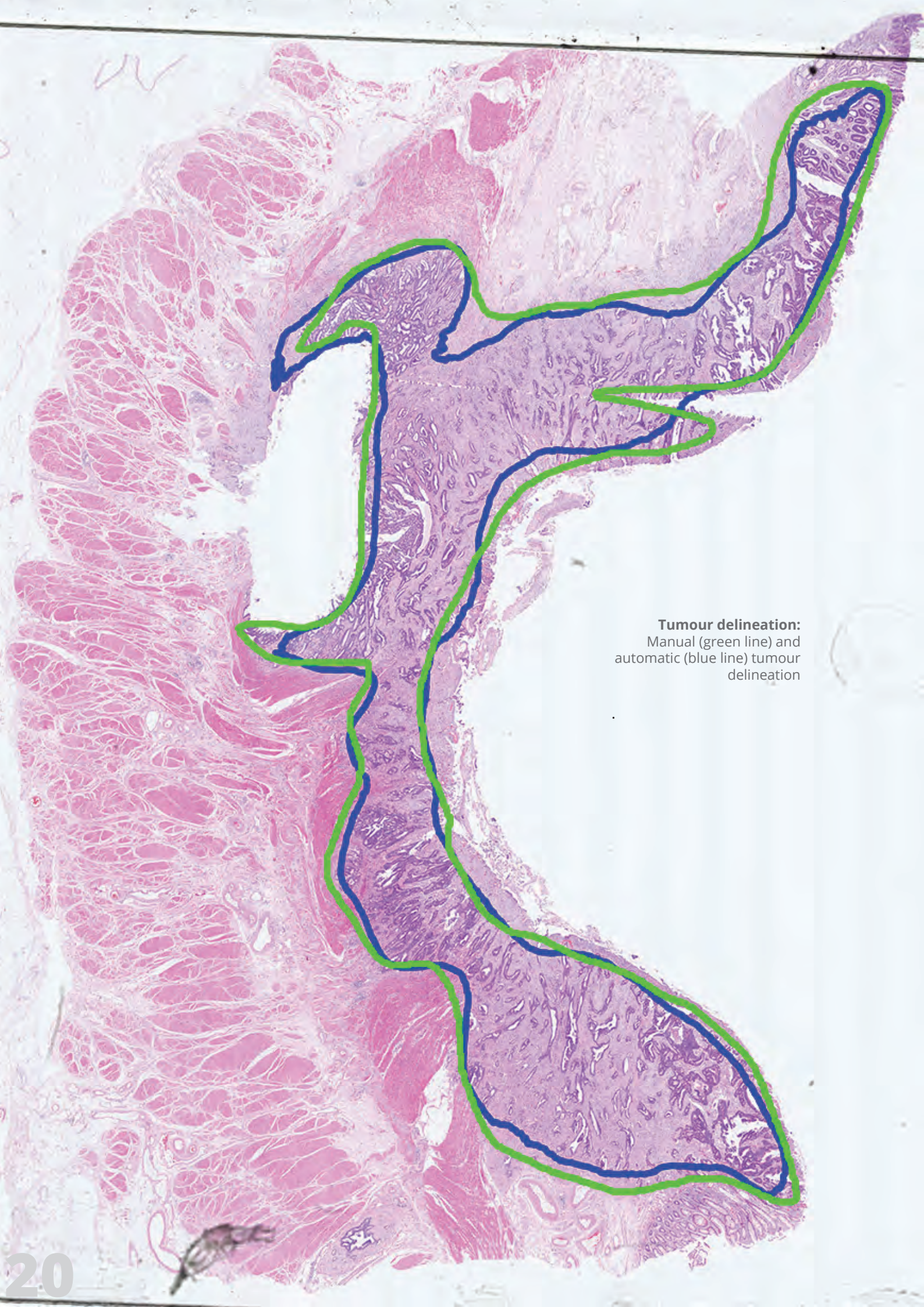
AUTOMATIC SEGMENTATION OF CELL NUCLEI

The detection and segmentation (draw around) of cell nuclei in Feulgen-stained histological sections is required for the subsequent analyses of DNA ploidy status and Nucleotyping. The Feulgen-stain binds specifically to DNA and allows the analysis of chromatin properties such as amount and structure. Visually the task of identifying and segmenting cell nuclei seems relatively simple to the human eye, but designing robust computer algorithms for this task has been a great challenge for digital image analysis.

In 2012 we published a method for the automated segmentation of cell nuclei in Feulgen-stained histological sections from prostate cancer *[ref Nielsen et al., Cytometry A. 2012 Jul;81(7):588-601. doi: 10.1002/cyto.a.22068]*. The method worked well in prostate cancer specimens but did not when applied to colorectal cancer specimens, where the cell- and tissue organisation is different. We have used deep learning with convolutional neural networks to develop a novel method for the automatic segmentation of cell nuclei in colorectal cancer. The new method is trained using 300 000 manually delineated cell nuclei in 139 patients and validated in histological sections from 51 independent patients, in which 104 000 manually delineated nuclei were available as a ground-truth in comparison. The image on the right illustrates result from the neural network model that was adapted to the problem. This new method has proven to work for prostate cancer as well, and a validation study in lung cancer is in progress.



Automatic segmentation of cell nuclei (fill with colour (burgundy) reflecting probability of true cell nucleus identification reported from the computer model, where deeper colours represent higher probabilities). Light blue outlines are manually segmented cell nuclei.



Tumour delineation:
Manual (green line) and automatic (blue line) tumour delineation

AUTOMATIC SEGMENTATION OF CANCEROUS TUMOUR REGIONS

Pathologists examine tumours and classify them as cancerous or non-cancerous. Our analyses of cancer patients are carried out in the cancerous tumour regions, and a method to identify these regions automatically is thus required. We have used deep learning with convolutional neural networks trained on a dataset of 2573 semi-manually delineated colorectal tumour samples to develop such a method. The resulting computer model is validated on a different set of 857 samples. The approach works well, with 93% sensitivity (the proportion of cancerous tumour region in the ground-truth also identified by the deep learning model) and 97% specificity (the proportion of non-cancerous region in the ground-truth also identified by the deep learning model), i.e. a good correlation with the pathologist's delineation.

DEEP NEURAL NETWORKS

A deep neural network is an artificial neural network with multiple layers between the input and output layers. The deep neural network finds the correct mathematical manipulation to turn the input into the output, whether it be a linear relationship or a non-linear relationship. The network moves through the layers calculating the probability of each output.

EVALUATION OF RESULTS

To evaluate the quality of our new methods to segment cell nuclei and tumour regions, we need methods that objectively quantify the degree of correspondence with the ground-truth. Ground-truth for the segmentation of cell nuclei is represented by manually segmented cell nuclei, while the ground-truth for cancerous tumour regions is in the form of annotations drawn by a pathologist. There is a lack of consensus for methods to evaluate the quality of these two approaches, although some methods are more frequently used than others, such as the Jaccard index which is defined as $\frac{|X \cap Y|}{|X \cup Y|}$ given two sets X and Y (i.e. the number of overlapping pixels in the predicted segmentation and the ground-truth divided by the total number of unique pixels in the predicted segmentation and the ground-truth). We have implemented a range of measures to evaluate our results robustly, exemplified for the automatic segmentation of cell nuclei below.

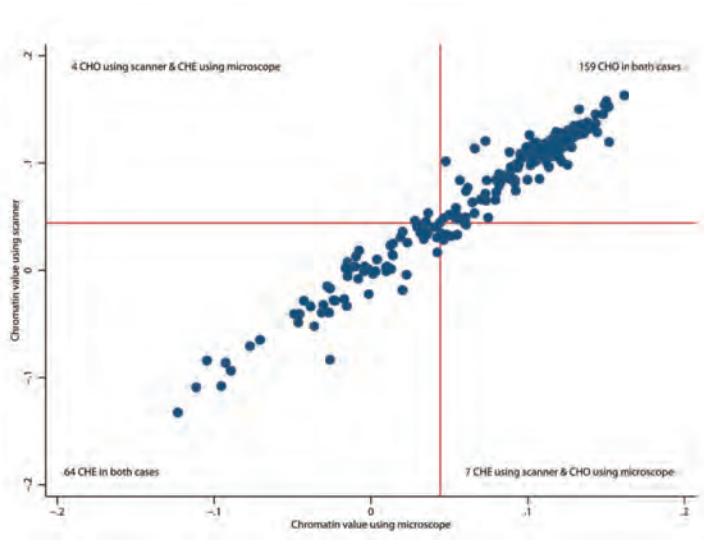
Measures	Above 0.2		Above 0.5		Above 0.8	
Counts	**Value**		**Value**		**Value**	
Cases	51		51		51	
Frames	3950		3952		3952	
Reference nuclei	104 863		104 890		104 890	
Reference without overlap	24 537		44 871		78 350	
Reference with overlap	80 326		60 019		26 540	
Proposed nuclei	104 616		66 774		26 212	
Proposed without overlap	26 358		8523		864	
Proposed with overlap	78 258		58 251		25 384	
Object detection	**Value**		**Value**		**Value**	
Sensitivity	0.838		0.709		0.389	
Precision	0.813		0.911		0.971	
Jaccard index	0.703		0.663		0.385	
F_1-score (Dice similarity)	0.836		0.798		0.555	
Object-level segmentation	**Mean**	**Median**	**Mean**	**Median**	**Mean**	**Median**
Sensitivity	0.886	0.928	0.902	0.931	0.913	0.938
Specificity	0.949	0.966	0.956	0.967	0.960	0.969
Precision	0.877	0.927	0.897	0.931	0.909	0.938
Rand accuracy	0.928	0.947	0.939	0.952	0.946	0.956
Jaccard index	0.790	0.845	0.824	0.859	0.848	0.876
Area under ROC curve	0.917	0.941	0.930	0.946	0.937	0.952
Gini ROC	0.734	0.782	0.766	0.799	0.792	0.819
F_1-score (Dice similarity)	0.869	0.916	0.894	0.924	0.908	0.934
Matthew's correlation	0.829	0.879	0.856	0.890	0.873	0.902
Bookmarker's informedness	0.834	0.882	0.858	0.892	0.874	0.904
Bookmarker's markedness	0.839	0.882	0.856	0.893	0.872	0.904
Normalized mutual information	0.646	0.689	0.679	0.708	0.707	0.732

Segmentation evaluation measures

WP4

DIGITAL PRECISION DIAGNOS-TICS

The DoMore! project aims to develop novel diagnostic and prognostic methods that are easily integrated into standard clinical routine, as well as a method to digitize and automate existing tasks in pathology.

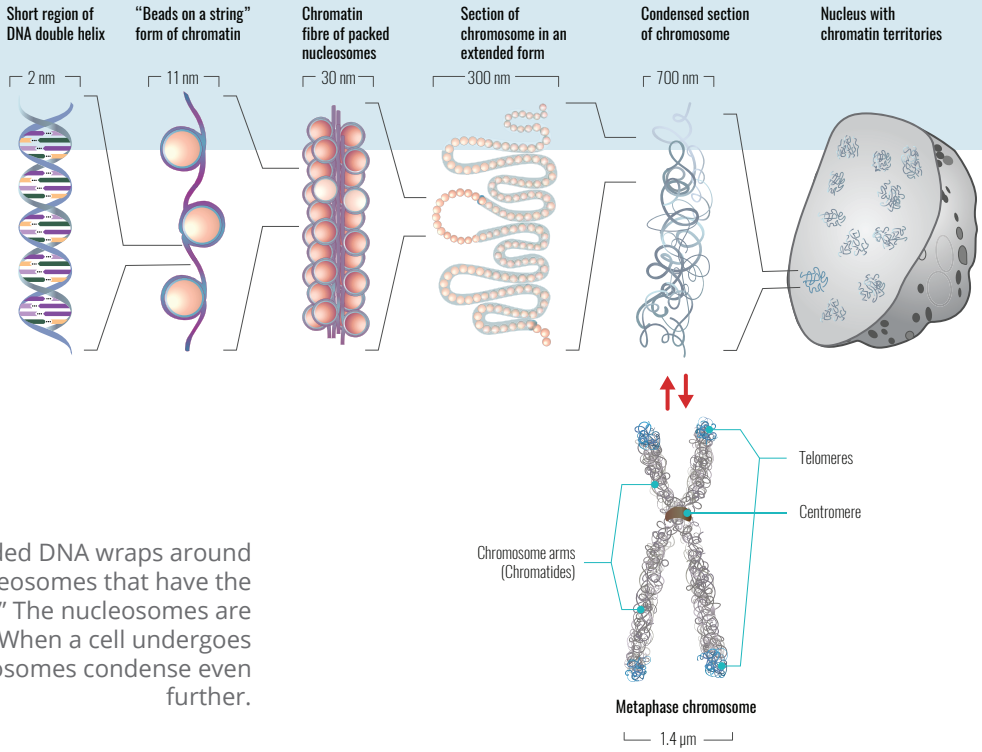


Scatter plot of chromatin value measured using a bright-field microscope and a whole-slide digital scanner. The red lines depict the threshold for dichotomising chromatin values; the classification is chromatin heterogeneous (CHE) if the chromatin value is smaller than the threshold and otherwise chromatin homogeneous (CHO). Pearson correlation coefficient was 0.98 (95% confidence interval [CI] 0.97-0.98; $p < 0.0001$) between the chromatin values and 0.89 (95% CI 0.86-0.91; $p < 0.0001$) between the chromatin classifications. Since the microscope was equipped with a 546 nm green filter and a monochrome digital camera while the scanner acquired colour images which were converted to grey scale by averaging, the integrated optical density (IOD) was typically far less in the scanner images and therefore the element width was reduced from 25 to 7.5 in the DNA ploidy histogram computed as a part of the image normalisation method, although the correlation was nearly as good without this adjustment (0.95 between chromatin values and 0.84 between chromatin classifications). From Kleppe & Danielsen, Oncotarget 9(65):32406-32407, 2018.

ABERRANT DNA CONTENT

in cancer cell nuclei is measured with DNA ploidy and is a marker for poor prognosis in several cancer types

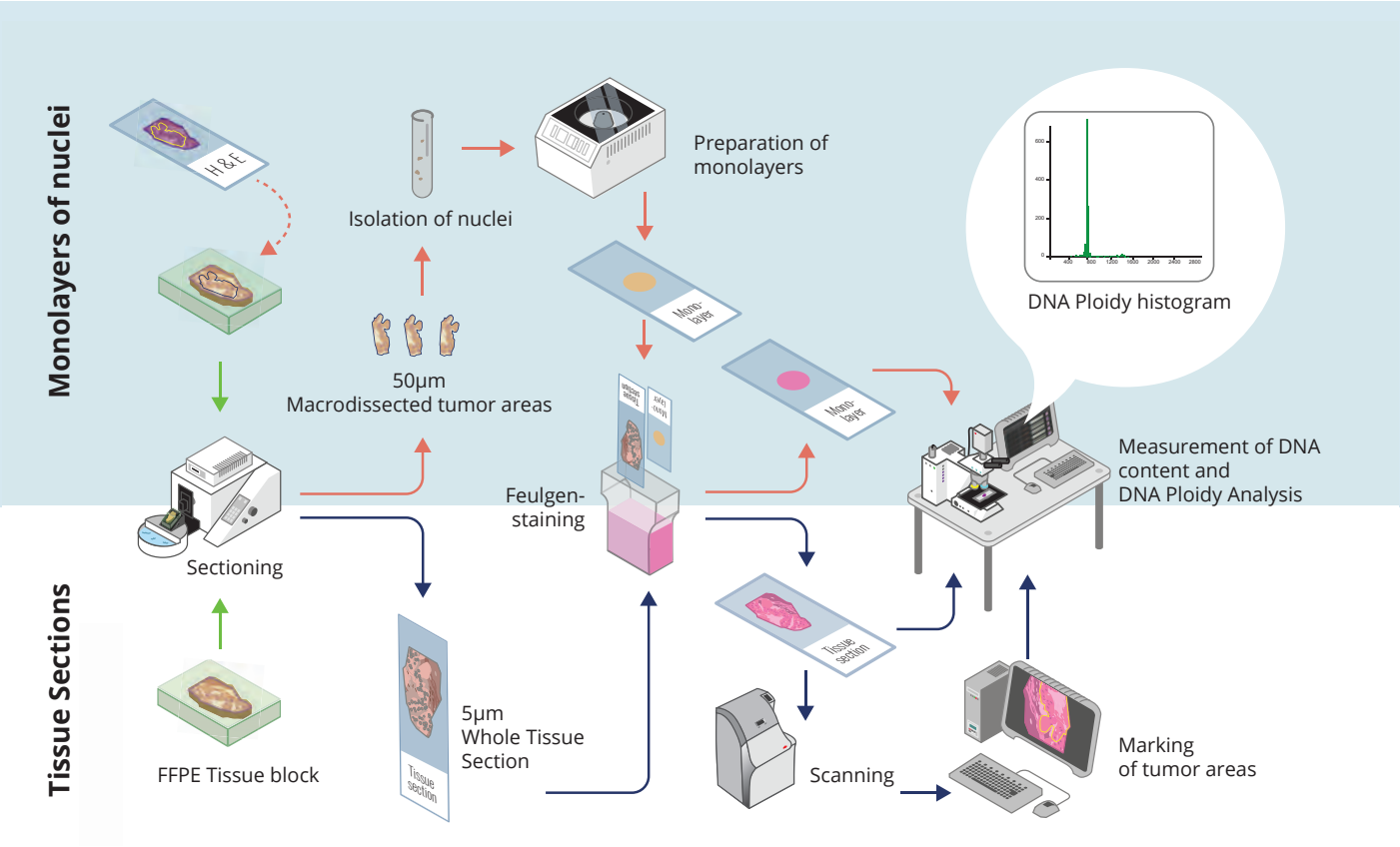
One main aim in the DoMore! project is to develop methods that are easily integrated into standard clinical routine. Our main methods DNA ploidy and Nucleotyping have been developed on specially prepared samples from the cell suspension, called monolayers, where intact cell nuclei are spun onto a glass slide and imaged with a microscope. Both the special preparation and the microscope are factors that decrease the availability of the methods. It is thus an aim to develop methods that work on routine histological sections scanned with a scanner rather than a microscope.



DNA organisation Double-stranded DNA wraps around histone proteins to form nucleosomes that have the appearance of "beads on a string." The nucleosomes are coiled into a 30-nm chromatin fiber. When a cell undergoes mitosis, the duplicated chromosomes condense even further.

DNA PLOIDY

Deviations from the normal configuration with two copies of each chromosome are relatively common in cancer cells and are associated with a worse prognosis for the patient. DNA ploidy by image cytometry estimates the total DNA content in cancer cell nuclei and classifies a sample as normal (diploid) or abnormal (non-diploid) based on the evaluation of a histogram of DNA content in about 1500 cell nuclei. The method is well established, and its prognostic impact in several cancer types is well documented *[ref Danielsen et al. Nat Rev Clin Oncol. 2016 May;13(5):291-304]*. In the DoMore! project we extend the DNA ploidy method first to make use of scanners rather than microscope, and then to be able to analyse the nuclei directly in routine sections, thus eliminating the need for special preparations of monolayers.

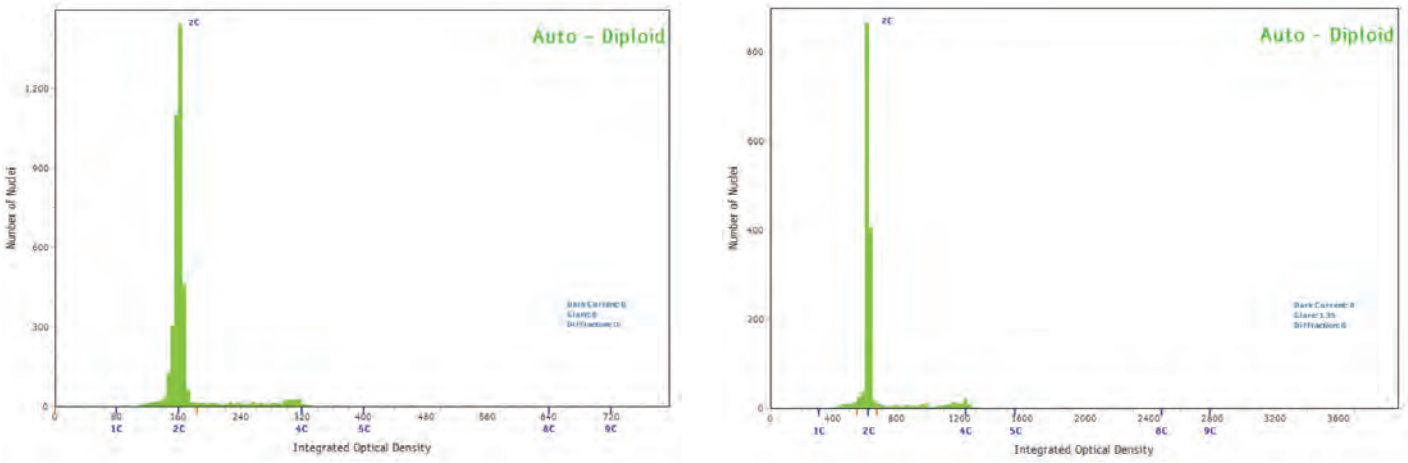


Alternative DNA ploidy pipelines in our laboratory

DNA PLOIDY ON A SCANNER

High-resolution scanners will become common equipment in pathology departments with the digitalisation of pathology. Scanning of slides is highly automated in these systems, with functions for scanning a rack of 400 or more slides overnight. We have adapted the DNA ploidy method on monolayers to work on the scanner. Due to the slightly lower resolution on the scanner compared to the microscope, the ability to detect small differences in DNA content is better on the microscope system, but for the majority of non-diploid samples, the DNA content representing the abnormal cell population is significantly higher than the normal cell population and as such not a problem for practical purposes. We have compared the resulting DNA ploidy classification in the two methods in 246 colorectal cancer samples where 236 (96%) had identical classification with the two methods. The few with different classification were due to cell populations with slightly aberrant DNA content compared to the normal diploid cell population where the scanner system did not identify them as aberrant.

The clinical implication of such hyperdiploid subpopolations is unclear and studies are underway to compare the prognostic power of these two different ways of analysing ploidy distributions in tumors.



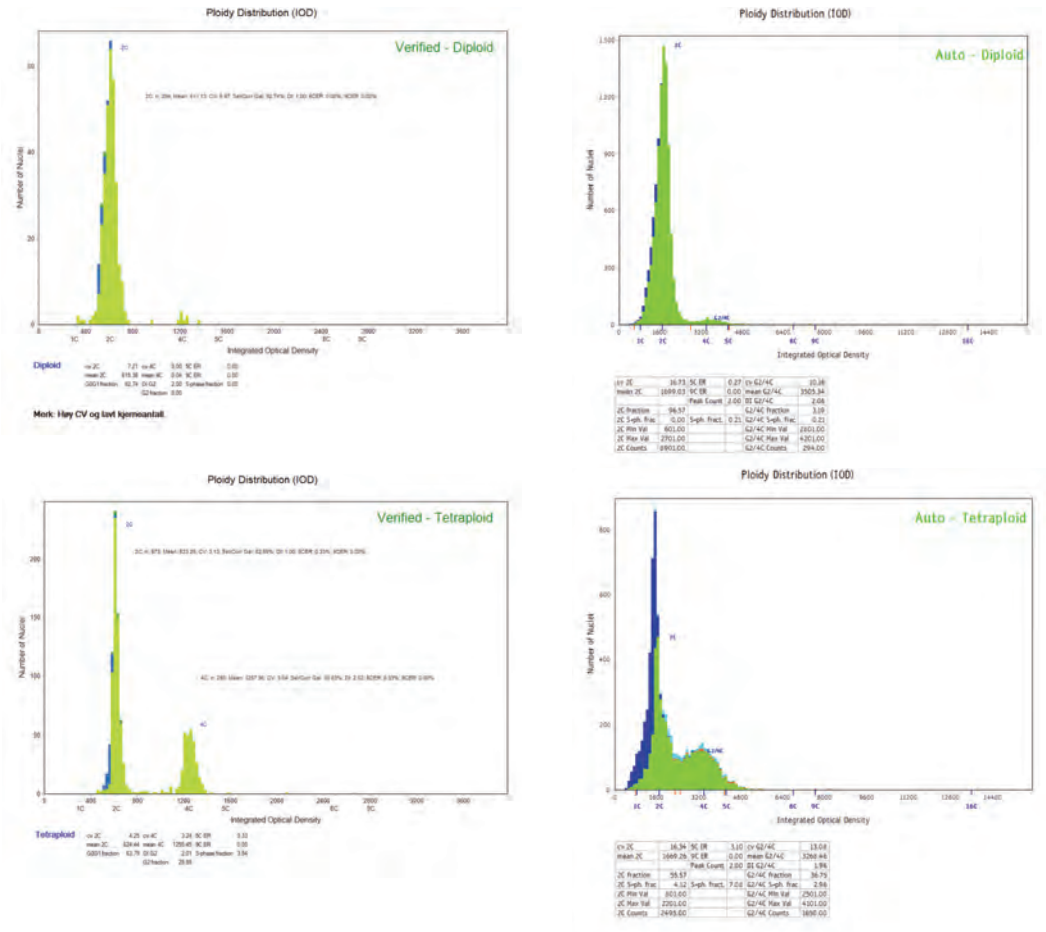
Above: DNA content histograms from the scanner (left) and microscope (right)

DNA PLOIDY IN HISTOLOGICAL SECTIONS

Methods that can be applied to routine sections are easier to implement in the clinic. The overall aim of this particular project is to develop a method for the estimation of DNA ploidy status in HE-stained routine sections. An important step in the development process is to estimate DNA ploidy status in Feulgen-stained histological sections; the HE-stain can be removed and replaced with Feulgen-stain and imaged before the section is restained with HE-stain. Feulgen-stain is DNA-specific and stoichiometric and thus allows for more precise measurement of DNA content than the HE-stain. The main challenges in the development of a method for the estimation of DNA content in histological sections are 1) the segmentation of cell nuclei and 2) the DNA content estimation from a profile (5µm) of a cell nucleus. A method for the automatic segmentation of cell nuclei has been developed based on deep learning with convolutional neural networks (see work package 3). Methods for the estimation of DNA content in thin histological sections have been reported earlier (e.g., Haroske et al. 1993) and we have implemented a variant of the method proposed by Haroske et al., and find good correspondence between the monolayer DNA ploidy estimate and the histological section estimate in tissues with near spherical shaped nuclei, such as prostate. In a dataset of 236 samples where we estimated DNA ploidy in both monolayers and histological sections, 188 (80%) had the same classification. So, the preliminary data shows promise, but further work is on-going to optimize and adapt the method.

	Diploid monolayer	Non-diploid monolayer
Diploid 2D	162	14
Non-diploid 2D	34	26

Correspondence between DNA ploidy classification in monolayer samples and histological sections



DNA ploidy histograms from two monolayer samples (left) and histological sections from the same tumour block (right)

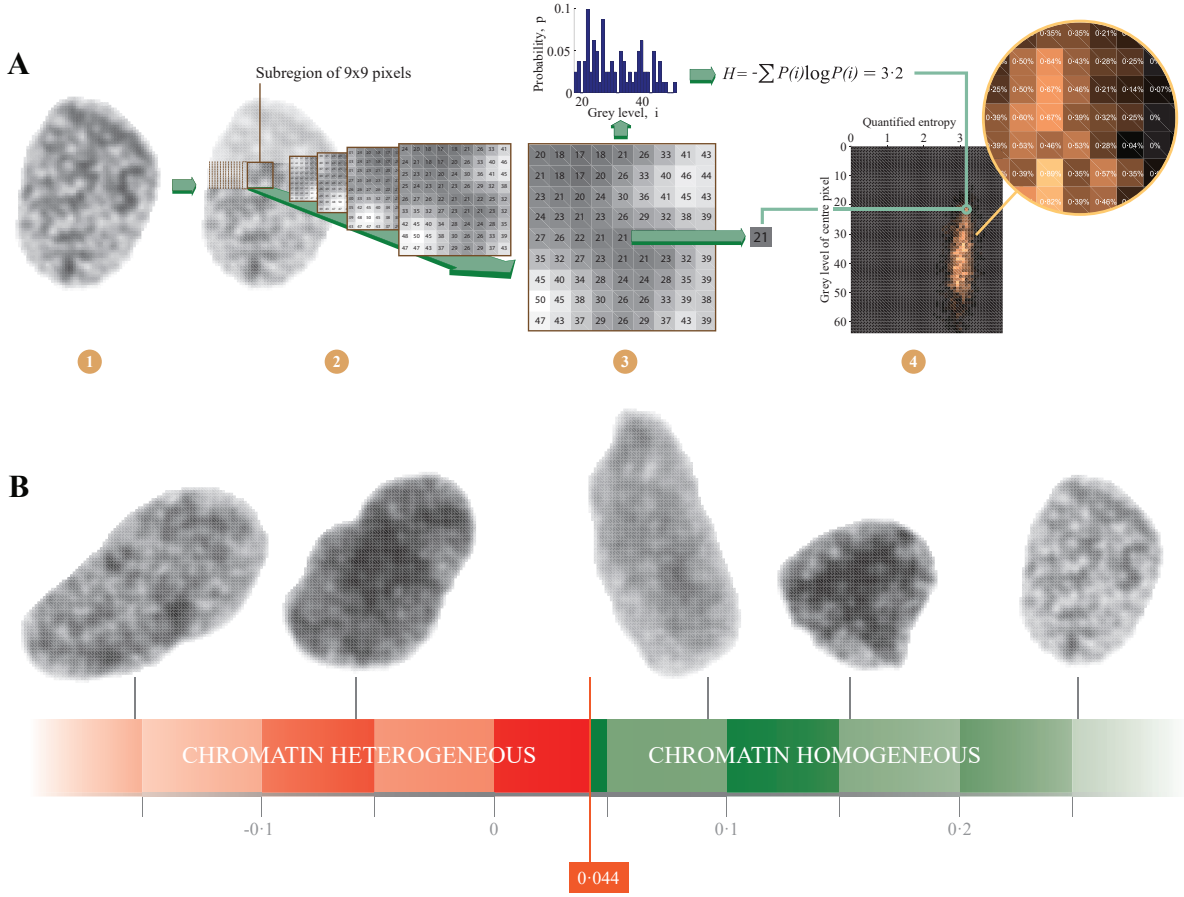
NUCLEOTYPING

The assessment of chromatin heterogeneity

NUCLEOTYPING

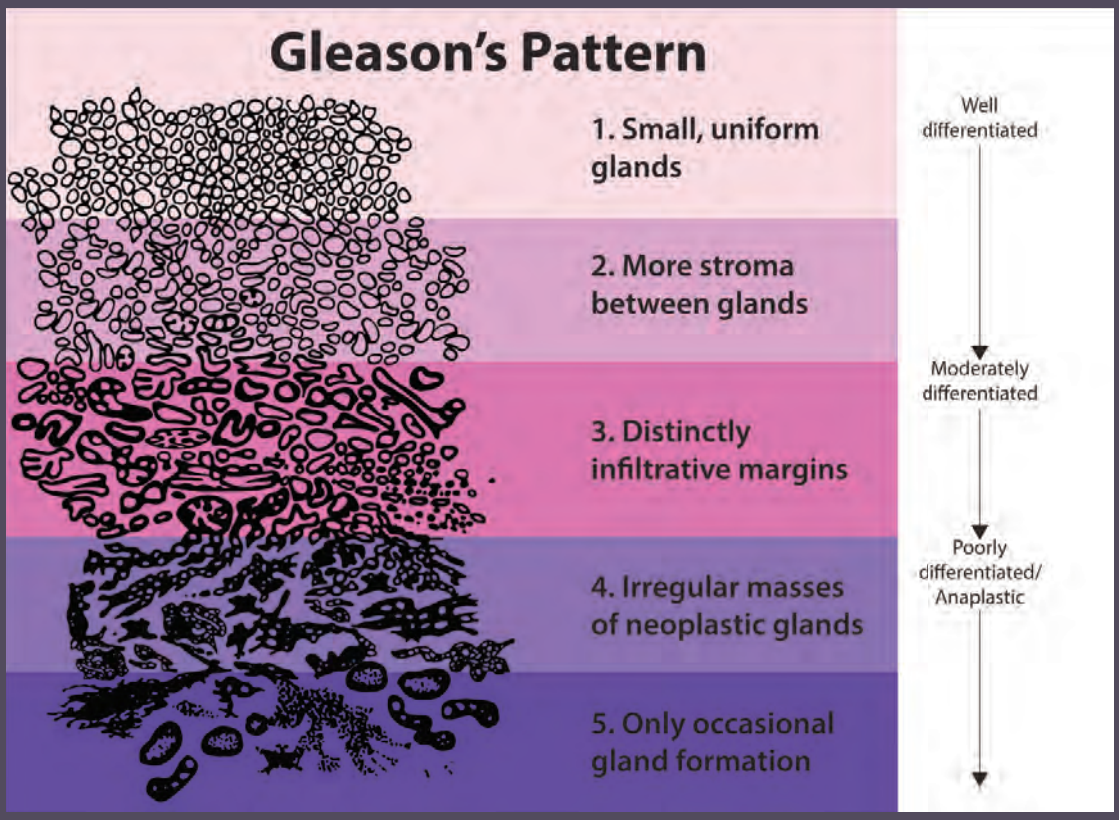
Chromatin structure in cancer cell nuclei is related to transcriptional activity and other cellular processes. We have developed a framework to characterize the chromatin structure based on grey level entropy in Feulgen-stained cell nuclei, where the disorder of grey levels reflects the DNA distribution in local regions. This characterization, termed Nucleotyping, has been developed on monolayer samples and found to discriminate cancer patients with different prognosis. The approach has been validated in datasets of different cancer types. In the DoMore!-project we have demonstrated that Nucleotyping is a general marker across cancer types *(Kleppe et al. Lancet Oncol. 2018 Mar;19(3):356-369)* and that the proportion of nuclei inhabiting variants of these properties also are strong prognostic markers across gynaecological cancer types *(Nielsen et al. J Natl Cancer Inst. 2018 Dec 1;110(12):1400-1408).*

We now focus our work on transferring the concept to histological sections, which are different from the specially prepared samples from cell suspension used till now for Nucleotyping. The cell nuclei are not complete due to the sectioning of 5µm sections and require more advanced methods to be automatically segmented. On the other hand, the cells are imaged in their original context in the tissue providing new opportunities for characterization. The previously described method for automated segmentation of cell nuclei in tissue sections with convolutional neural networks is also used in this project. We currently work on the colorectal cancer dataset (about 600 patients treated at Aker University Hospital) in which the automated segmentation method is developed as well as another colorectal cancer dataset (The Gloucester Colorectal Cancer Study with more than 900 patients) to develop the Nucleotyping method in histological sections. Till now we have implemented the methods used for Nucleotyping in monolayers and will evaluate their prognostic impact. Preliminary results are promising with trends indicating a prognostic role for the same methods as in monolayer samples. Extensions of the existing method include incorporating the contextual information from the surrounding tissue as well as the implementation of novel methods for the quantification of chromatin properties. This wil be a focus project in DoMore! for the next year or two.

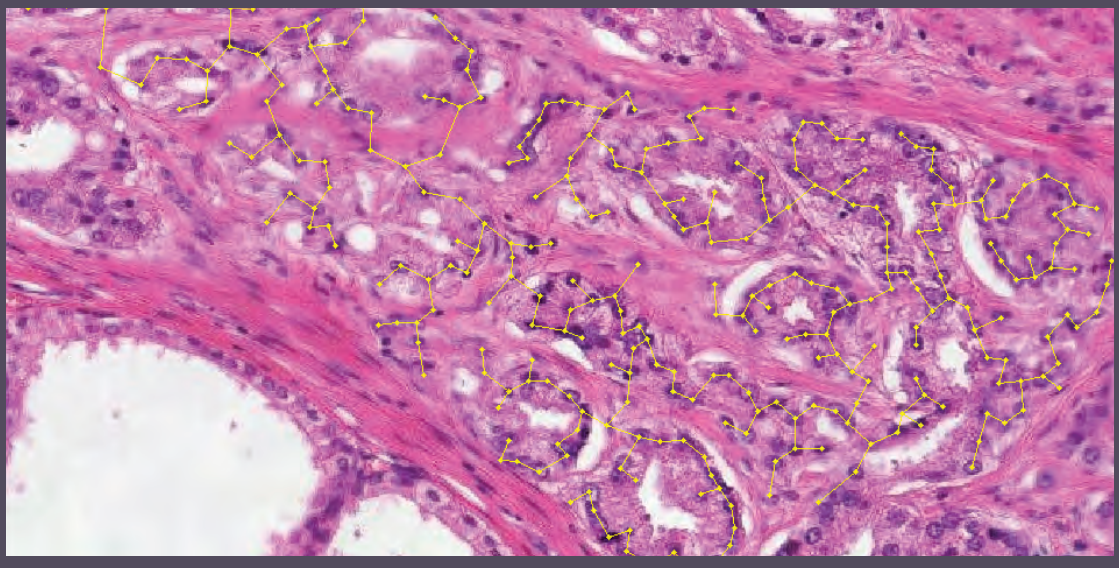


Computation of the grey level entropy matrix (GLEM) and visualisation of nuclear images

(A) Illustration of GLEM computation. (1) A nuclear image. (2) Each nuclear pixel is taken to be the centre of a square subregion, here with a side length of nine pixels. (3) For each subregion, two quantities are extracted (the grey level of the centre pixel [here 21] and the entropy of the grey levels in the subregion [here 3·2]); the entropy H is a variability characteristic of the probability mass function P(i) (ie, the histogram that gives the probability P that grey level i occurs in the subregion). (4) The two quantities extracted from the subregion will together identify a position in a two-way table. The table cell position corresponding to the subregion in figure part 3 of panel A is marked by a green circle in part 4 of panel A. The occurrence is counted by incrementing the value at the table cell position (initially, all table cell values are 0), and the computation of the two quantities and incrementation of the corresponding table cell value is performed for every subregion of the nuclear image. The resulting table describes the frequency of each pair of centre grey level and surrounding entropy and is normalised by its total count to provide the bivariate probability mass function called the GLEM. The two-way table visualised in part A4 is the GLEM of the nuclear image in part A1. (B) Depiction of five nuclear images and their chromatin value. The threshold applied to dichotomise the chromatin value was 0·044.



The Gleason grading system



Example of detected cell nuclei and the corresponding minimum spanning tree (yellow dots and lines)

AUTOMATIC GLEASON SCORING

Donald Gleason was an American physician and pathologist who identified tissue patterns in prostate cancer specimens that were associated with patient prognosis. A grade between 1 and 5 was assigned each of the tissue patterns, such that a higher grade corresponded with more aggressive disease. The dominant and predominant grades were identified, resulting in a Gleason score representing the two main tissue patterns, e.g., 3+4(=7). The Gleason grading system assessed by uropathologists is a very strong prognostic marker for prostate cancer patients and is routinely assessed in this patient group. Specialized pathologists are a scarce resource and increasingly so with the rising incidence numbers for this cancer type.

Furthermore, the intra- and interobserver variation for Gleason scoring is significant. There is thus a need for alternative methods for the assessment of Gleason score in prostate cancer tissue specimens. We have developed a method that estimates the Gleason score in HE-stained tissue sections from prostate cancer specimens based on cell organisation and duct features. The method identifies cell nuclei and ducts, calculates a minimum spanning tree based on the cell nuclei positions and uses features from the minimum spanning tree together with features describing the ducts in a support vector machine classifier to estimate a Gleason grade to the image. The method has been trained and validated on a set of images with homogeneous Gleason grade assessed by uropathologists. The validation results are good, and we are currently working on implementing a version of the application that has a simple user interface, and that can be applied to whole sections.

Discovery Data Set CCR = 90%

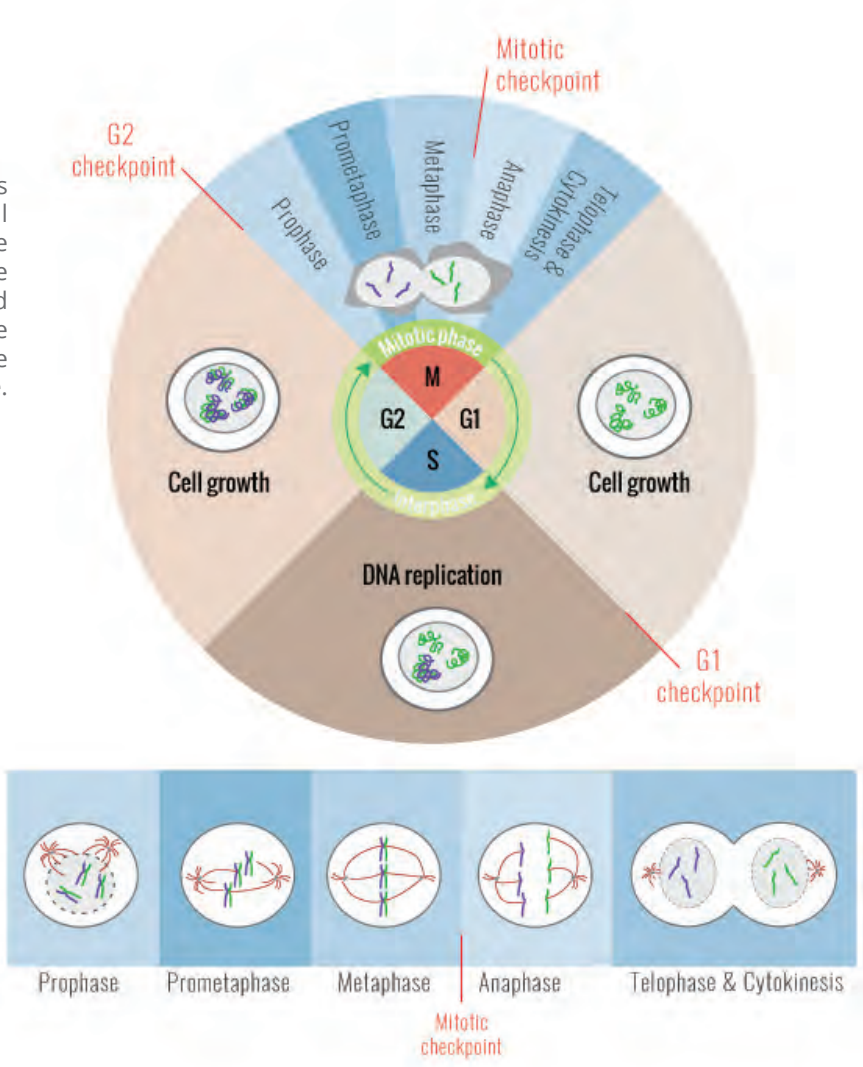
	Gleason grade 3	Gleason grade 4
Gleason grade 3	172	28
Gleason grade 4	19	240

Validation Data Set CCR = 84%

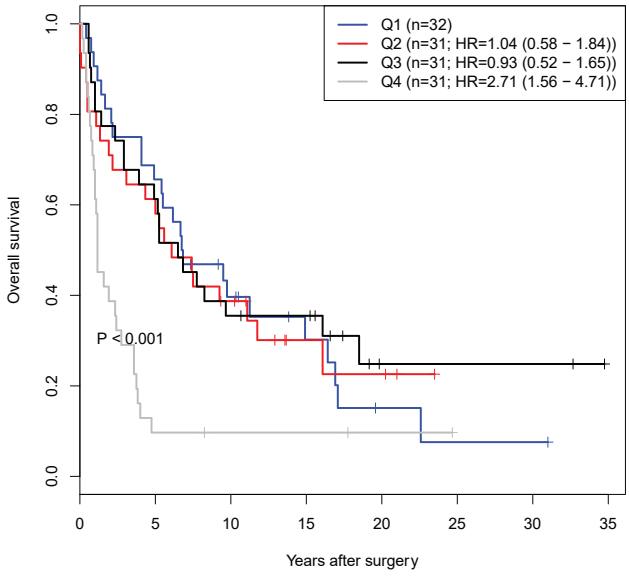
	Gleason grade 3	Gleason grade 4
Gleason grade 3	52	14
Gleason grade 4	10	77

Concordance between pathologist (columns) and computer (rows) based Gleason grading in the discovery (training) and validation (test) datasets.

The cell cycle describes the series of stages eukaryote cells pass through during cell division. This description is essential in the understanding of carcinogenesis and for the development of new methods for improved characterisation of cancer. Several of the DoMore! projects, such as DNA ploidy, are closely linked to the cell cycle.



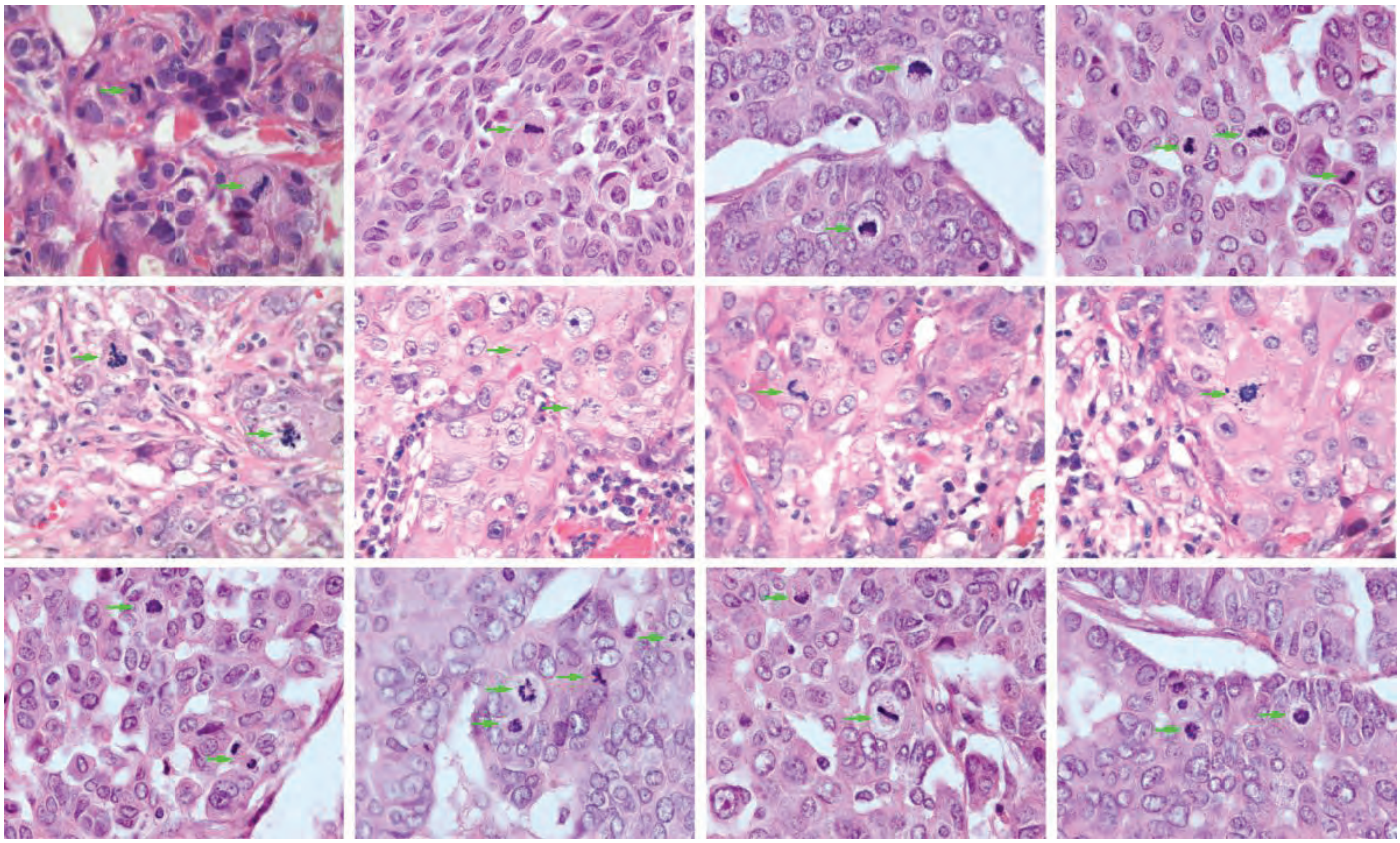
Kaplan-Meier plot of survival probability grouped by quartiles of standardized computerized mitotic figure counts in 125 leiomyosarcoma patients



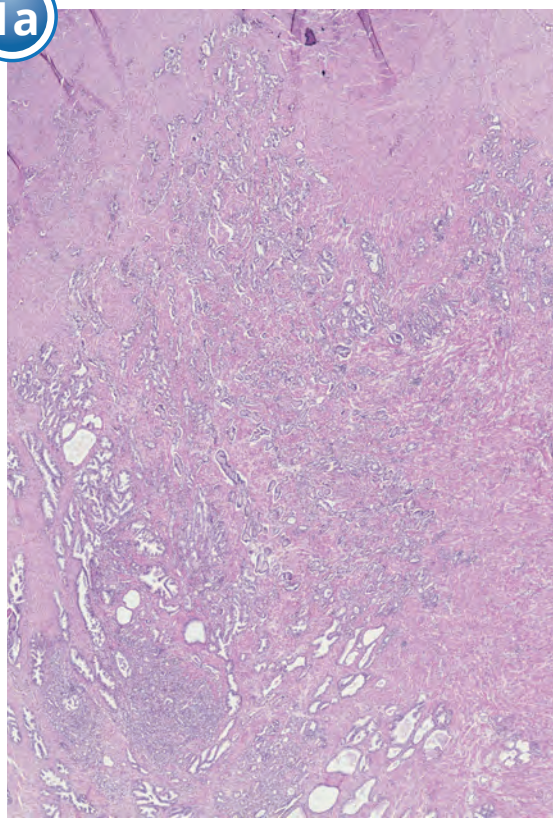
AUTOMATIC MITOTIC INDEX ASSESSMENT

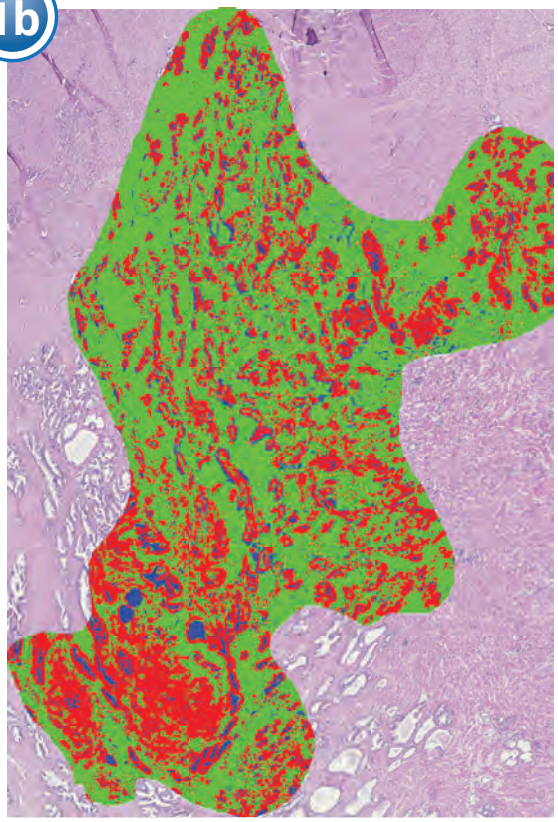
The proportion of cells undergoing cell division is a prognostic marker for several cancer types, where a higher proportion for most cancer types is associated with a worse prognosis. The counting of cells undergoing mitosis is one method to assess the degree of cell proliferation in a tissue sample. Pathologists normally count the number of mitotic figures in ten high-power fields, providing a standardized count for the patient. The process is laborious, and the inter- and intraobserver variation is significant. There is thus the interest in an accurate and computerized measurement of this property. We have trained a convolutional neural network to identify mitotic figures in a publicly available breast cancer dataset used in challenges in biomedical imaging such as the Tumor Proliferation Assessment Challenge (TUPAC). The resulting neural network model was then validated on a set of leiomyosarcomas and indicated that the method works as intended. We continue our work with improving the neural network model and extending the validation dataset to comprise all uterine sarcomas in Norway between 1970 and 2000 as described by Abeler et al. *(ref Abeler et al. Histopathology. 2009 Feb;54(3):355-64)*.

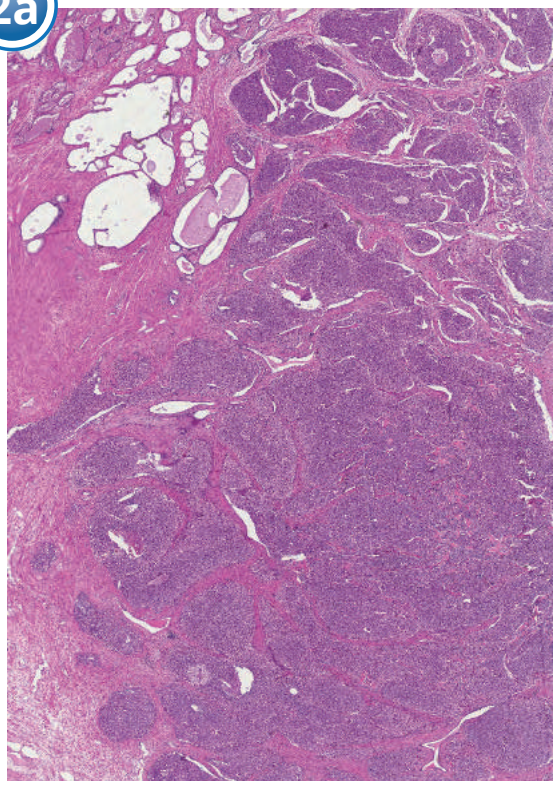
When completed, the plan is to validate this method on a large breast cancer material to demonstrate clinical feasibility and implementation.

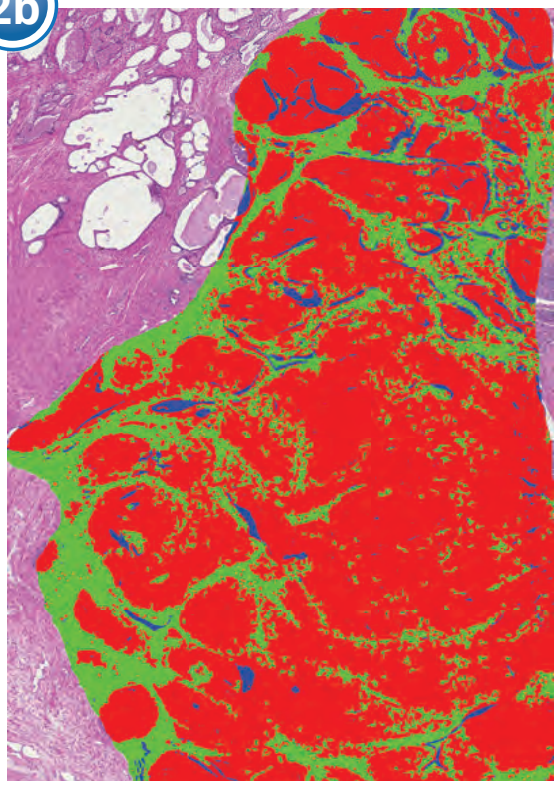


Examples of mitotic index annotation in breast cancer tissue samples





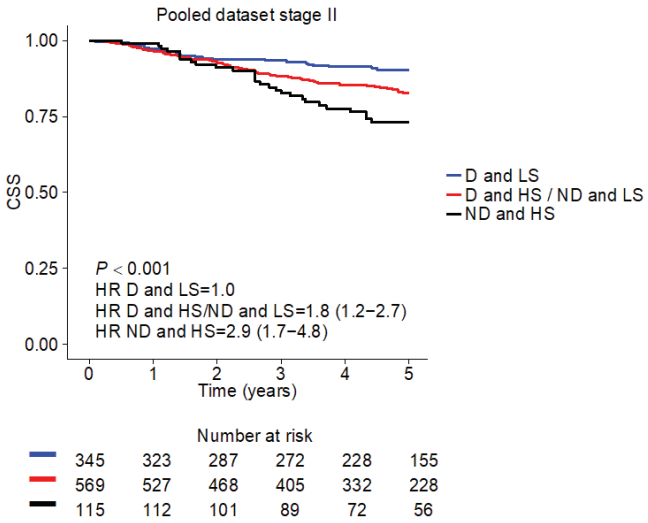




Stroma detection
Stroma cells are automatically detected and shown in red.

AUTOMATED ESTIMATION OF STROMA FRACTION

The tumor microenvironment is the cellular environment in which a tumor exists. A tumor and the surrounding microenvironment are closely related and interact constantly. For simplicity, the tumor is often analyzed in isolation although the dependence on the surrounding tumor environment is well known. Stroma is a major component in the tumor microenvironment, and the stroma-to-epithelial proportion has been reported to be a prognostic factor for colorectal cancer patients based on manual assessment *[e.g., Huijbers et al. Ann Oncol. 2013 Jan;24(1):179-85]*, where a higher stroma fraction is associated with a worse prognosis. We have developed a method to assess the stroma fraction in HE-stained routine histological sections automatically and validated its prognostic impact in colorectal *[Danielsen et al., Ann Oncol. 2018 Mar 1;29(3):616-623]* and prostate cancer *[Ersvær et al. manuscript submitted]*. We use a fixed threshold to categorize the stroma fraction as low or high. Furthermore, we have combined the stroma fraction estimate with DNA ploidy status to integrate prognostic information from a tumor and its microenvironment. The method can be easily integrated into a clinical routine at a low cost.



Kaplan-Meier plot of cancer-specific survival probability grouped by ploidy and stroma status in colorectal stage II patients.

WP5

DATA FLOW AND QUALITY

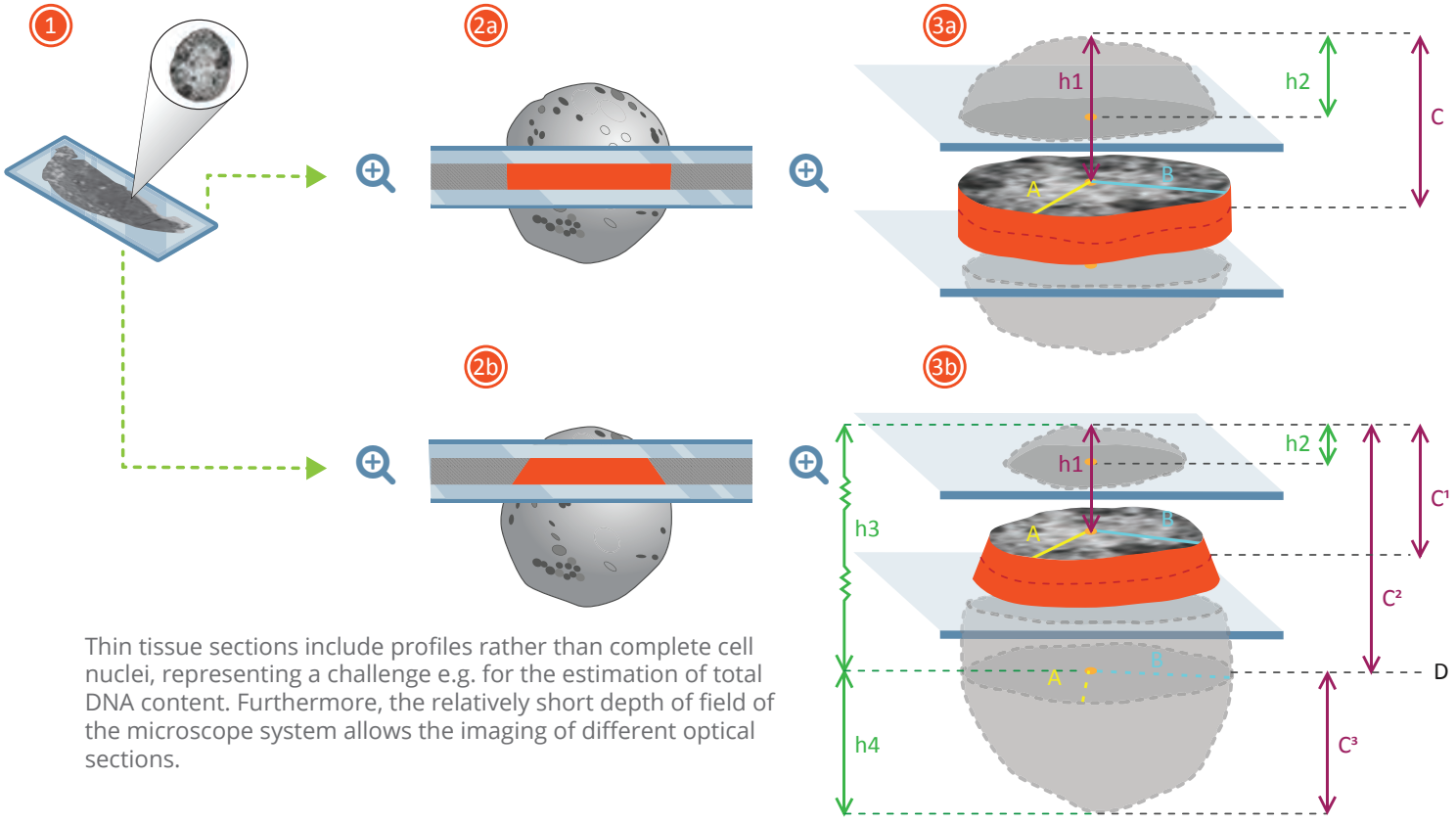
The DoMore! project handles a huge number of images. The success of the project relies on these images to have good enough quality to ensure that the algorithms works. Our priority now is on understanding and control the various factors that influence the image quality of the scanner, such as magnification, resolution, image compression, darkcurrent, glare, diffraction and not at least focus. A main challenge is to understand and mimick how a human distinguishes a focused from an unfocused image.

The scanner company is responsible for producing images with reliable and stable quality, but experience has taught us that scanners more or less frequently produces images that are partially or totally out of focus. Manual control detects many of these cases. However, an automatic algorithm could increase the likelihood of detecting unfocused images and thus ensure a better result in the final analysis. In Task 5.1 and 5.2, focusing quality is the common denominator. The effort has therefore been to establish an automatic grading of the focus quality of images produced by microscopes and scanners. Two different approaches have been used.

Wp5 consists of seven subtasks. So far, the activity has been focused on Task 5.1; A survey of image-focusing algorithms with emphasis on texture analysis, Task 5.2; Analysis if image quality variation and Task 5.3; File formats and compression.

MICROSCOPY IMAGES

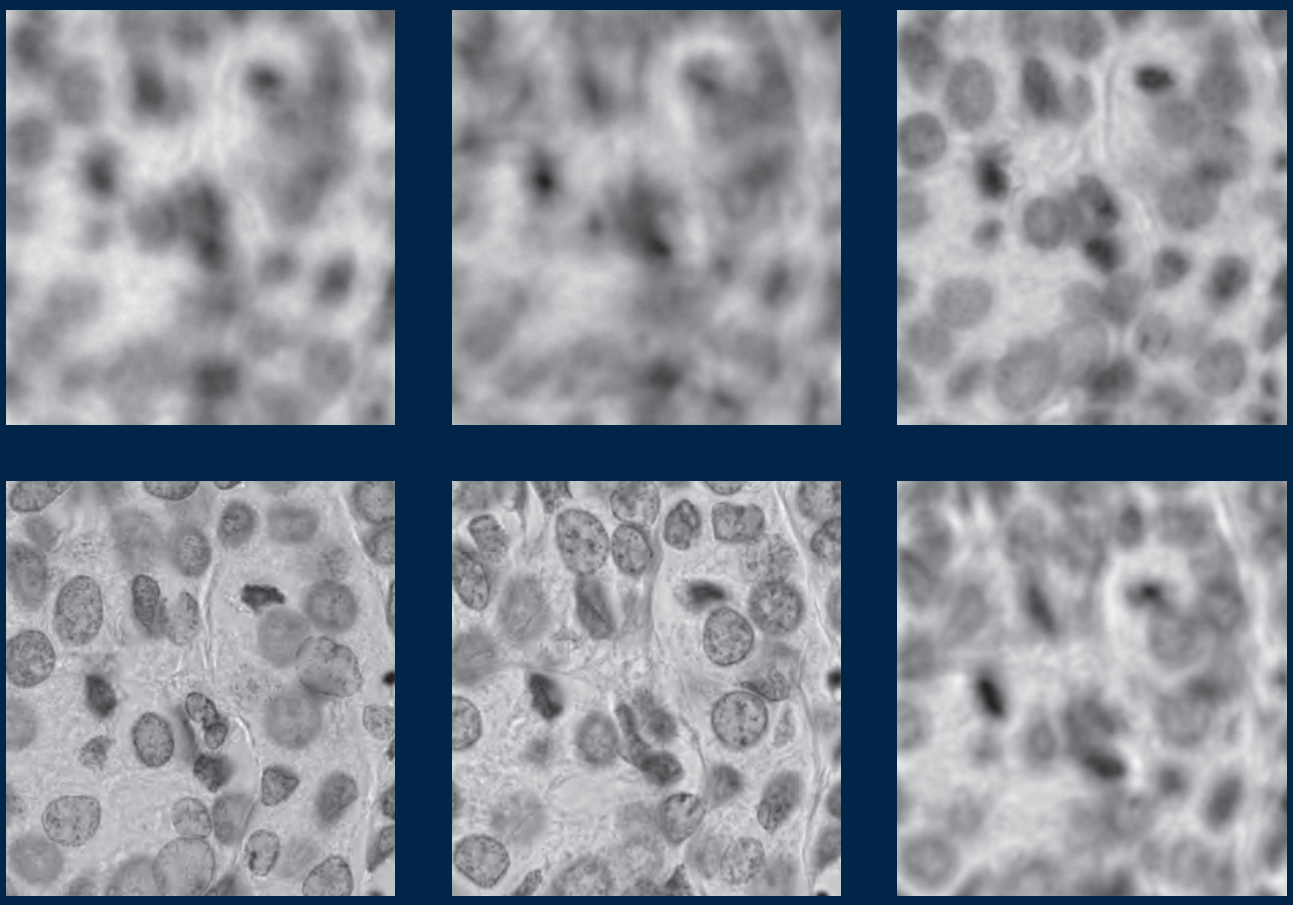
For microscopy imaging, we do have the possibility of adjusting the image focus, i.e., performing autofocus. In autofocusing, a quick search is done to find the lens position that gives the best-focused image according to some criteria. This can be done sequentially for a small increment in the lens position, or iteratively for decreasing increments, moving the lens back and forth to the position where a given focus measure of the given image is a maximum. Even though autofocusing has been a long-standing topic, most of the papers published on the subject are devoted to proposing a new method that is marginally different from previous ones and testing the performance on just a few selected images. A few surveys and comparisons of focus measures exist, but the focusing problem is generally related to the imaging modality and field of application.



Thin tissue sections include profiles rather than complete cell nuclei, representing a challenge e.g. for the estimation of total DNA content. Furthermore, the relatively short depth of field of the microscope system allows the imaging of different optical sections.

Our task is to find an optimally focused image of a microscopy image of cells, containing structures at several scales having a variety of grey level gradients, where the final image processing step is a statistical and/or structural texture analysis of the grey level morphology of the cell nuclei. Thus, we do not only need to find the "best" focusing algorithm, but also to investigate if the choice of focusing algorithm can result in some selection effects when it comes to the subsequent texture analysis.

To determine the performance of various focusing metrics, a thorough literature study was performed. The first draft of a report on different focusing algorithms suitable for cell nuclei microscopy imaging has been completed. Based on the report, a pilot study has been carried out. In the study, five of the most promising focusing algorithms were implemented and tested on sets of microscopy images where for the same sample, the focus had been varied in a controlled way for a large number of possible focus depths. The preliminary results show that several of the metrics performed well, i.e., identified a focus depth that co-aligned with the one picked by a trained user, but that the Energy of Laplacian might be the most promising one. To continue this study, new data needed to be recorded.



Finding the optimal focus of each scans has proven to be a challenge.

SCANNED SLIDES

In DoMore!, images from more than 60,000 slides are scanned. Ideally, all of these images are of high quality having high contrast and a perfect focus, as these images form the foundation of further analysis. Based on the knowledge gained in Task 5.1 we wanted to develop a tool to monitor the image quality. The scanner produces an image, and the task was to determine if this image was unfocused. A trained human can reliably establish this in a couple of seconds. Having a computer program doing the same is more challenging.

A larger dataset was collected to evaluate various metrics and ideas. It consisted of pairs of 90 scans collected on three different scanners. Each pair consisted of an unfocused scan (detected manually), and the same scan rescanned and controlled to have an acceptable focus quality. From each scanner (Aperio, Hamamatsu XR and Hamamatsu NZ), ten scans from three different tumor type (prostate, colorectal and lung) were collected. The scans were manually annotated to identify areas of different tissue type (tumor, connective/muscle/fatty tissue, necrotic areas). The annotation was performed with up to 20 times zoom level.

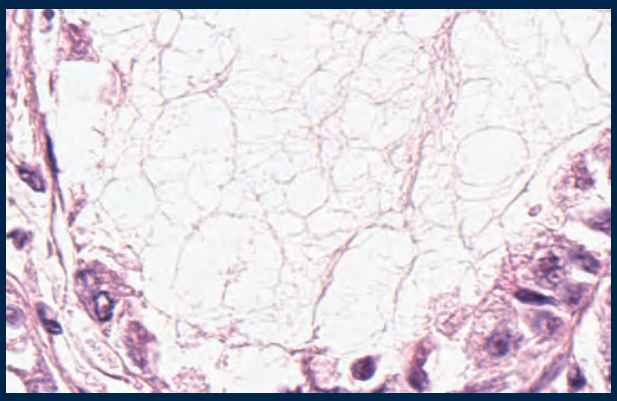
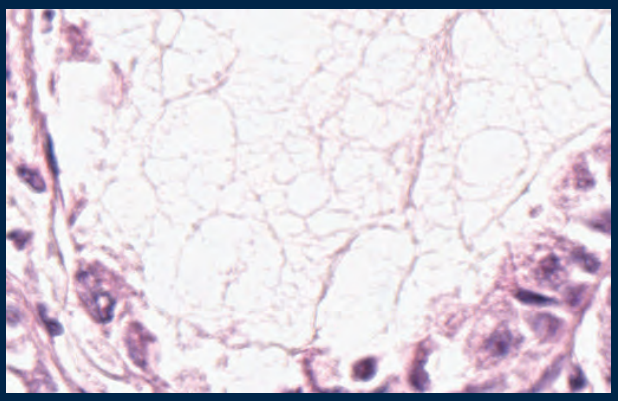
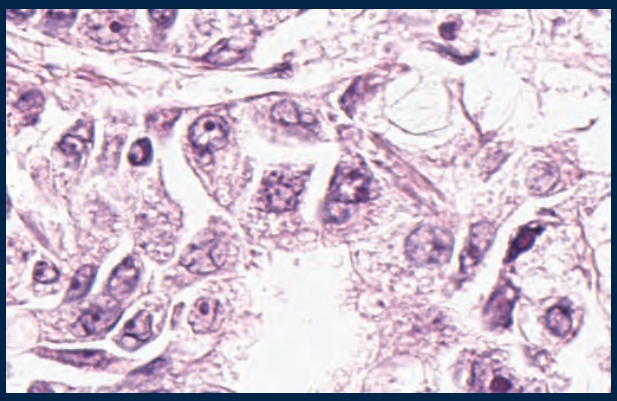
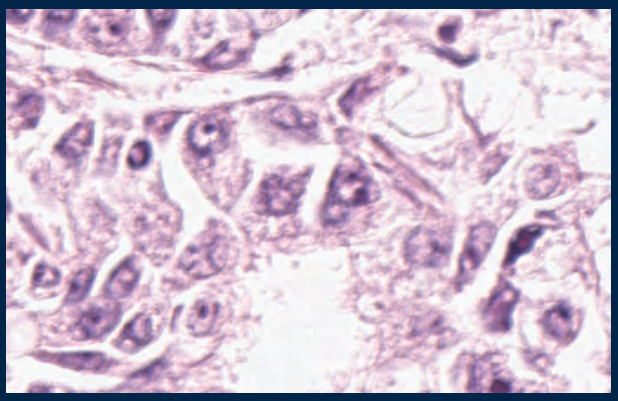
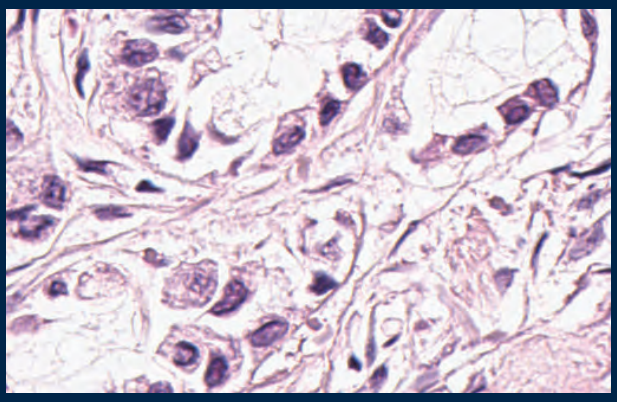
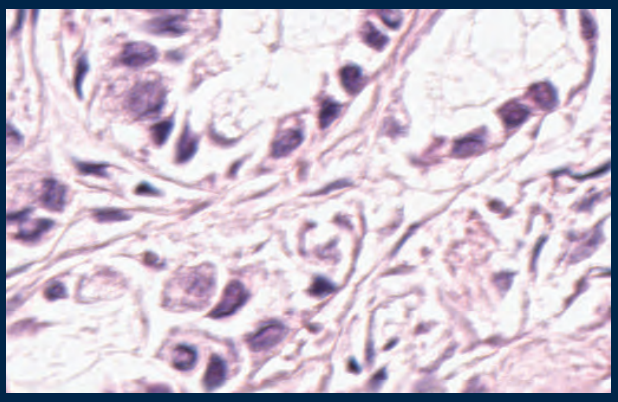
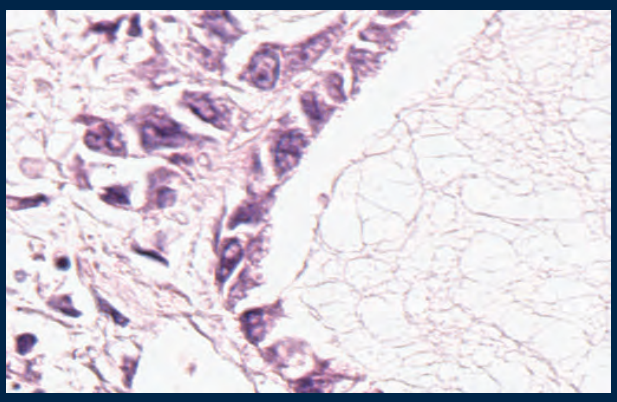
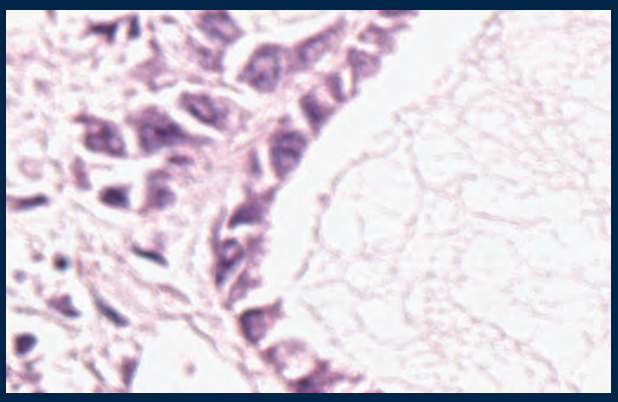
The current status of the project is that, so far, no metric is found that can separate all incidences of focused from unfocused images. Our priority now is on understanding and mimicking how a human distinguishes a focused from an unfocused image.

The wanted outcome of this project is an algorithm that determines if a given scan is of a quality good enough to be included in the following analysis, or if the slide should be rescanned or removed from the analysis. The algorithm should work for both full scans and for tiles, and ideally be independent of scanner and tissue type. We have a good indication that it is possible to establish such an algorithm for a given scanner and tissue type. If it will be independent of scanner and tissue type, remains to be proven.

In the following, we will continue working on a global algorithm for scanned slides. When this is established, we will return to microscopy images and take up again our project on auto-focusing algorithms. We will then investigate the relationship between focusing measures and texture features.

Another important task will be to closely examine all the other parameters that influence image quality with the goal to reduce the coefficient of variation of integrated optical density meassured in scanner images, which is 1.5 times higher than for microscope images.

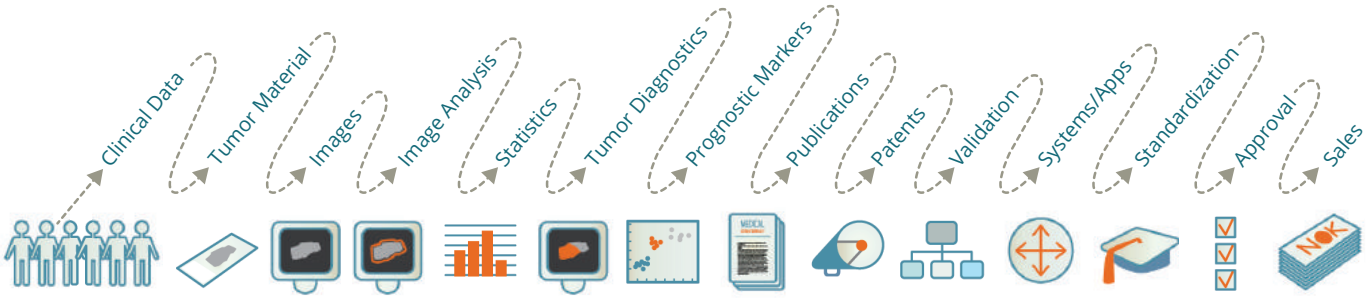
This work package will be assigned more resources in 2019 - 2020, as they are becoming available after completing tasks in other work packages.

WP6

COMMERCIALI-SATION, HEALTH ECONOMY AND PROGNOSTIC DECISION SUPPORT

As the development activities of DoMore! has progressed and the potential products have advanced in the pipeline, the activities connected to the commercialisation of the projects inventions started during autumn 2018.



The main goal for the first phase has been to outline the general strategies for the transfer of technology, rights, and knowhow to a commercial company as well as the principles of ownership of such a company and the revenue streams back to the stakeholders.

OVERALL STRATEGIC COMMERCIALISATION GOAL

DoMore!s commercialisation strategy is included in the project's overall strategy. We believe it will add its value to the project, making it more than a passive owner and seller of IP rights and software components. The base of this strategy is the understanding that in silico pathology rests on the foundation of digital pathology. A "passive" strategy is more vulnerable to the dynamics of the industry, such as a slow distribution of digital pathology. Therefore, a viable strategy includes options for wider involvement in the pathology value chains. Although the business strategy is at a very early stage, it is vital that the commercial setup is designed to facilitate this.

PRODUCTS AND SERVICES

A range of artificial intelligence products resulting from the research and development has been identified through the commercialisation discussions so far. The products that are already in the pipeline can be roughly divided into three main groups:

- Pathology workflow optimization
- Prognostic markers
- Screening tools

These three types of products fit in different parts of the value chain of the industry and are thus believed to require different commercial approaches. For some of the products, simple license agreements with hardware vendors may be the optimal commercialisation strategy, whereas for others the optimal strategy for the stock owners may be to develop services or enter into strategic collaborations with existing service providers. Also, planned products within clinical decision support, which is not yet in the pipeline, may require yet another commercialisation strategy.

TRANSFER MODEL

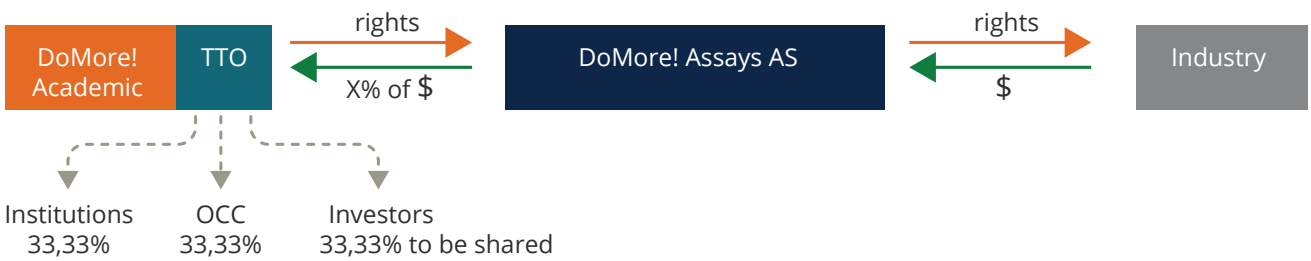
The strategy for IP-transfer assumes the creation of a limited company (AS =aksjeselskap in Norwegian) that acquires the necessary technology, rights/ licenses and know-how from the project. The transfer model (to the right) illustrates the flow of IP and money from the DoMore! -project to the new company, which for this strategy is named DoMore! Assays AS. In the process of setting up the company, the project is represented by the TTO (Technology Transfer Organisation), which in this project is Oslo Cancer Cluster, while the yet empty company is represented and set up by Radforsk.

PATENTS

DoMore scientists have developed machine learning algorithm trained on some scanner images. Each image is divided into tiles which are then used to train the deep learning algorithm. The trained algorithm may then be used to evaluate images.

A UK patent application covering the methodology described above was filed in the UK on November 16, 2017. A PCT application was filed on November 9, 2018.

TRANSFER MODEL



INITIAL FUNDING OF THE COMMERCIAL COMPANY

Radforsk and OCC will, on behalf of the parties negotiate the commercial terms, i.e., license/royalty agreements for the first products. OCC as TTO will develop an investment memorandum, describing the investment opportunity. Many business ideas and concepts based on the developed products have been outlined during the work package activities, which will be described in the memorandum. The company will be based on this, secure funding for the first part of the commercialisation through a private placement, facilitated by Radforsk.

THE INITIAL COMPANY SET UP

The main element of the company setup is the search for an appointment of CEO which will formally lead the commercialisation activities. The primary task for the CEO will be to develop a business plan and execute this plan to initiate a revenue stream.
The CEO will from the beginning need at least one technical resource (developer) to be able to deliver the technology to potential customers. He or she will also need an advisory board consisting of people with deep technical knowledge of the products and the potential markets.

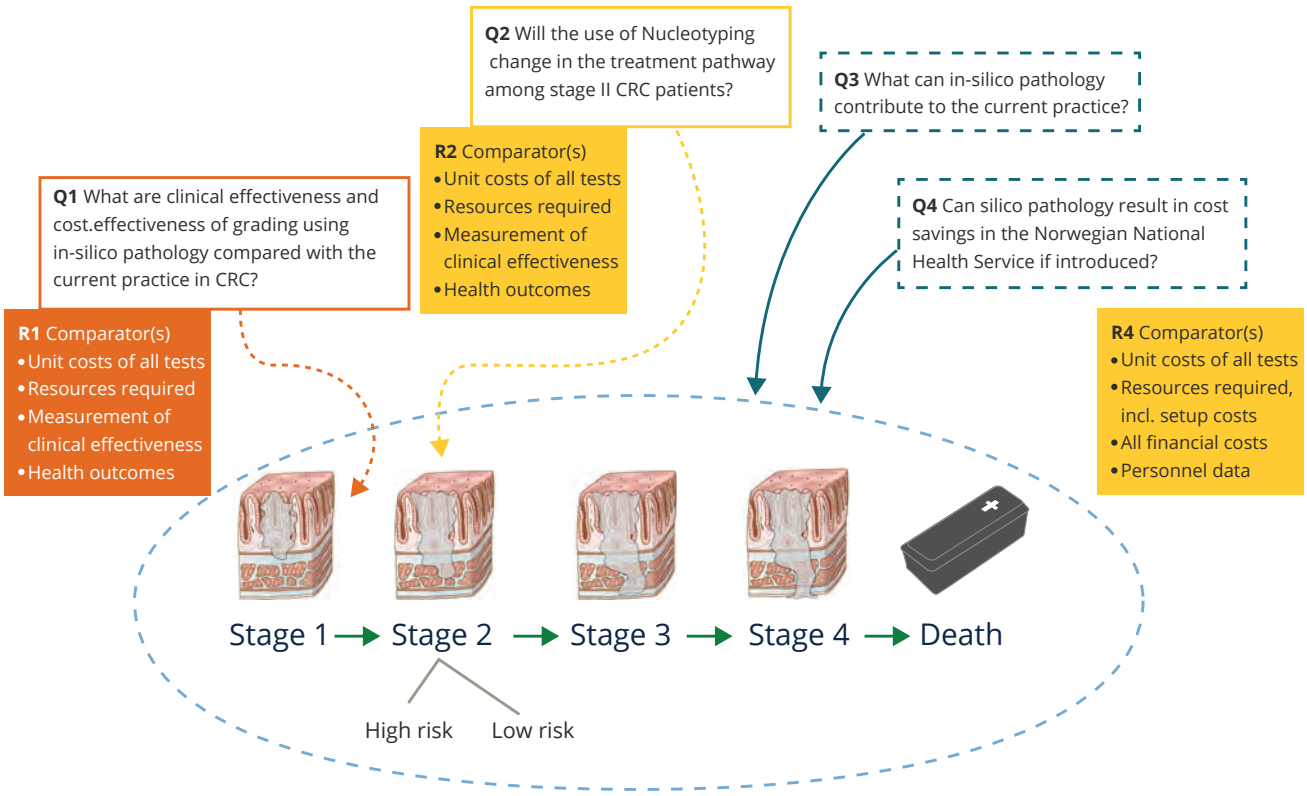
As the first products are seen as ready for commercialisation now, the goal for the CEO should be to be able to show a cash stream during the second year of operation.

CLINICAL DECISION SUPPORT SYSTEM IN COLLABORATION WITH DIPS ASA

As a partner in DoMore! and as a provider of Electronic Patient Journal (EPJ) to the majority of Norwegian hospitals, DIPS is committed to developing an integrated Clinical Decision Support System.
The team has started a process based on different deep learning systems to analyze and score prognostic information from DoMore!. The result will be integrated with scores from radiology imaging and clinical information.

A TEST LAB

DIPS Arena will set up a test lab that allows computer scientists from DoMore! and DIPS to work in collaboration. The collaboration structures have been designed Q4 in 2018 and the emphasis in the first phase, Q1, and Q2 of 2019 will be on technology integration. The last year of the DoMore! project will be geared towards testing and presentation of findings in a clinical dashboard- solution - the WP.



Summary of key questions/areas for discussion

HEALTH ECONOMICS

There is limited literature on Health economics and cancer biomarkers. This is because it is challenging to conduct a periodic review of evidence and update of existing clinical guidance based on the best evidence. For example, the drugs Cetuximab and Panitumumab that use companion diagnostics have been evolved through further refinement of technologies leading to the identification of subgroups who would be benefitted from the technologies of interest (NICE, 2017).

Contrary to the rising expectations, there has been less enthusiasm about the research and development of biomarkers and personalized medicines in relation to the return of investments (ROI) for technology developers and the budget impact for healthcare payers. It is because the assessment of the pharmaceutical and companion diagnostic package can be undertaken in much the same way as for pharmaceuticals without companion diagnostics. However, in circumstances where alternative tests are available, for example, proprietary test kits or "in-house tests" for the same biomarker that would fulfil the requirements of the pharmaceutical marketing authorization, the amount of extra effort to fully evaluate these alternative options is likely to exceed the available resources and timeframe in technology appraisals (NICE 2013).

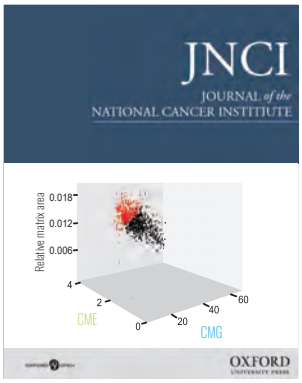
Defining the decision problem is the first step to be taken for the health economic evidence directly related to diagnostic biomarkers (Payne 2014; NICE 2011). This has been the main task so far. The following steps are to conceptualize the model: Select the relevant decision analytic model, build the decision analytic model, collect data to populate the decision analytic model, calculate expected costs and consequences, identify uncertainty and present the results.

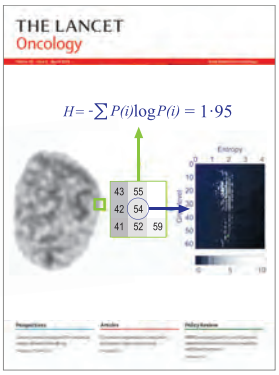
The complete process is rather resource intensive. Based on the current evidence base of DoMore! project, Nucleotyping or Histotyping could be seen as one of two potential areas for the scoping (decision problems); "Q2 clinical and cost-effectiveness of Nucleotyping or Histotyping among stage II CRC patients for adjuvant therapy" (previous figure). The availability of good quality data to populate a Health Economics analysis.

WP7

VALIDATION AND DISSEMINATION

Test and validation studies, scientific publications and dissemination are all included and a strategic part of the DoMore! project.





VALIDATION

When working with machine learning and deep learning, a proper predefined experimental design is crucial to ensure reliable results that are replicable in other labs. The experimental design for most of our projects is based on three main steps; learning, testing and validation, where each step usually employ independent data. Independence is of particular importance in the final validation studies. Ideally, these should be performed in an external lab, but this is in most cases not practically possible, and there are unfortunately very few replication studies published in science.

The Histotyping method (WP2) has been tested on a dataset produced externally, to ensure that the method is invariant to technical artefacts from the preparation of images. We have used multiple scanners to ensure invariance to technical artefacts in any given scanner. The final validation of Histotyping in colorectal cancers will be performed on tissue sections from more than 1000 patients with either stage II or stage III disease from a well described international clinical trial (Quasar 2).

For the methods developed in WP3, tumour delineation and nuclear segmentation, a design with an independent validation study is challenging because an objective ground truth is not available. It is in practice only possible to objectively assess to what extent the automatic segmentation equals the subjective segmentation defined by trained personnel. The final validation of each method in WP3 will therefore be empirically based on the validation studies where the methods are used, i.e., if Histotyping is successfully validated, so is the tumour delineation, as it was used to define the tumour areas to be analysed in Histotyping.

In WP4 we have several very different methods and employ slightly different experimental designs, but again use the three-step setup learning-testing-validation. For ploidy/stroma and Nucleotyping, the validation is explained and discussed in detail in the published papers (see next page).

The automated Gleason grading method, which is based on the established grading system developed by Gleason, is evaluated by comparing with results from several specialised uropathologists. The main challenge here is that there is a relatively high inter-observer variation. Our method do very well when we compare to those cases where the pathologists are in agreement, but one would have to argue that those cases might be the easier ones to grade. The final validation study will therefore be based on the actual outcome of the patients in a new cohort. The challenge in doing so is partly the tumour heterogeneity and partly that the clinical outcome of a prostate cancer patient is not known until 10-15 years after treatment. The validation cohort for this project is yet not defined.

For mitotic index, we have used a public material for learning, and the trained model has been independently tested in a cohort of uterine sarcomas, where we also have manual mitotic counting for comparison. The test is however on the outcome of the patient. We are currently seeking an external material for the final validation, which might be a breast cancer cohort, as this is a cancer type where mitotic index is already implemented in clinical routine.

SCIENTIFIC PUBLICATIONS

We have so far published 9 papers from the DoMore! project and another 8 related papers are published by DoMore! partners during these first half of the project. Our studies have been well received by biomedical journals with a mean impact of 12.4, which we find very satisfactory.

A further 5 manuscripts have recently been submitted for publication.

DoMore!

1. **Clinical utility of chromatin analysis.**
   Kleppe A, Danielsen HE.
   Oncotarget. 2018 Aug 21;9(65):32406-32407.
   PMID: 30197751

2. **Mutation burden and other molecular markers of prognosis in colorectal cancer treated with curative intent: results from the QUASAR 2 clinical trial and an Australian community-based series.**
   Domingo E, Camps C, Kaisaki PJ, Parsons MJ, Mouradov D, Pentony MM, Makino S, Palmieri M, Ward RL, Hawkins NJ, Gibbs P, Askautrud H, Oukrif D, Wang H, Wood J, Tomlinson E, Bark Y, Kaur K, Johnstone EC, Palles C, Church DN, Novelli M, Danielsen HE, Sherlock J, Kerr D, Kerr R, Sieber O, Taylor JC, Tomlinson I.
   Lancet Gastroenterol Hepatol. 2018 Sep;3(9):635-643.
   PMID: 30042065

3. **Association Between Proportion of Nuclei With High Chromatin Entropy and Prognosis in Gynecological Cancers.**
   Nielsen B, Kleppe A, Hveem TS, Pradhan M, Syvertsen RA, Nesheim JA, Kristensen GB, Trovik J, Kerr DJ, Albregtsen F, Danielsen HE.
   J Natl Cancer Inst. 2018 Dec 1;110(12):1400-1408.
   PMID: 29684152

4. **Chromatin organisation and cancer prognosis: a pan-cancer study.**
   Kleppe A, Albregtsen F, Vlatkovic L, Pradhan M, Nielsen B, Hveem TS, Askautrud HA, Kristensen GB, Nesbakken A, Trovik J, Wæhre H, Tomlinson I, Shepherd NA, Novelli M, Kerr DJ, Danielsen HE.
   Lancet Oncol. 2018 Mar;19(3):356-369.
   PMID: 29402700

5. **Prognostic markers for colorectal cancer: estimating ploidy and stroma.**
   Danielsen HE, Hveem TS, Domingo E, Pradhan M, Kleppe A, Syvertsen RA, Kostolomov I, Nesheim JA, Askautrud HA, Nesbakken A, Lothe RA, Svindland A, Shepherd N, Novelli M, Johnstone E, Tomlinson I, Kerr R, Kerr DJ.
   Ann Oncol. 2018 Mar 1;29(3):616-623.
   PMID: 29293881

6. **Computer-aided diagnostics in digital pathology.**
   Bengtsson E, Danielsen H, Treanor D, Gurcan MN, MacAulay C, Molnár B.
   Cytometry A. 2017 Jun;91(6):551-554.
   PMID: 28640523

7. **Tumour heterogeneity poses a significant challenge to cancer biomarker research.**
   Cyll K, Ersvær E, Vlatkovic L, Pradhan M, Kildal W, Avranden Kjær M, Kleppe A, Hveem TS, Carlsen B, Gill S, Löffeler S, Haug ES, Wæhre H, Sooriakumaran P, Danielsen HE.
   Br J Cancer. 2017 Jul 25;117(3):367-375
   PMID: 28618431

8. **BCL9L Dysfunction Impairs Caspase-2 Expression Permitting Aneuploidy Tolerance in Colorectal Cancer.**
   López-García C, Sansregret L, Domingo E, McGranahan N, Hobor S, Birkbak NJ, Horswell S, Grönroos E, Favero F, Rowan AJ, Matthews N, Begum S, Phillimore B, Burrell R, Oukrif D, Spencer-Dene B, Kovac M, Stamp G, Stewart A, Danielsen H, Novelli M, Tomlinson I, Swanton C.
   Cancer Cell. 2017 Jan 9;31(1):79-93.
   PMID: 28073006

9. **Cancer Prognosis Defined by the Combined Analysis of 8q, PTEN and ERG.**
   Silva MP, Barros-Silva JD, Ersvær E, Kildal W, Hveem TS, Pradhan M, Vieira J, Teixeira MR, Danielsen HE.
   Transl Oncol. 2016 Dec;9(6):575-582.
   PMID: 27916292

RELATED PUBLICATIONS AUTHORED BY DOMORE! PARTNERS

10. **The value of additional bevacizumab in patients with high-risk stroma-high colon cancer. A study within the QUASAR2 trial, an open-label randomized phase 3 trial.**
Huijbers A, van Pelt GW, Kerr RS, Johnstone EC, Tollenaar RAEM, Kerr DJ, Mesker WE.
J Surg Oncol. 2018 Apr;117(5):1043-1048.
PMID: 29448309

11. **Universal screening for Lynch syndrome in a large consecutive cohort of Chinese colorectal cancer patients: High prevalence and unique molecular features.**
Jiang W, Cai MY, Li SY, Bei JX, Wang F, Hampel H, Ling YH, Frayling IM, Sinicrope FA, Rodriguez-Bigas MA, Dignam JJ, Kerr DJ, Rosell R, Mao M, Li JB, Guo YM, Wu XY, Kong LH, Tang JH, Wu XD, Li CF, Chen JR, Ou QJ, Ye MZ, Guo FM, Han P, Wang QW, Wan DS, Li L, Xu RH, Pan ZZ, Ding PR; Written on behalf of AME Colorectal Cancer Cooperative Group.
Int J Cancer. 2018 Dec 6. [Epub ahead of print]
PMID: 30521064

12. **Critical research gaps and recommendations to inform research prioritisation for more effective prevention and improved outcomes in colorectal cancer.**
Lawler M, Alsina D, Adams RA, Anderson AS, Brown G, Fearnhead NS, Fenwick SW, Halloran SP, Hochhauser D, Hull MA, Koelzer VH, McNair AGK, Monahan KJ, Näthke I, Norton C, Novelli MR, Steele RJC, Thomas AL, Wilde LM, Wilson RH, Tomlinson I; Bowel Cancer UK Critical Research Gaps in Colorectal Cancer Initiative.
Gut. 2018 Jan;67(1):179-193.
PMID: 29233930

13. **Assessment of DNA Ploidy in the ProMisE molecular subgroups of endometrial cancer.**
Proctor L, Pradhan M, Leung S, Cheng A, Lee CH, Soslow RA, Gilks CB, Talhouk A, McAlpine JM, Danielsen HE, Hoang LN.
Gynecol Oncol. 2017 Sep;146(3):596-602.
PMID: 28647100

14. **Changes in Chromatin Structure in Curettage Specimens Identifies High-Risk Patients in Endometrial Cancer.**
Hveem TS, Njølstad TS, Nielsen B, Syvertsen RA, Nesheim JA, Kjæreng ML, Kildal W, Pradhan M, Marcickiewicz J, Tingulstad S, Staff AC, Haugland HK, Eraker R, Oddenes K, Rokne JA, Tjugum J, Lode MS, Amant F, Werner HM, Bjørge L, Albregtsen F, Liestøl K, Salvesen HB, Trovik J, Danielsen HE; ENITEC network/MoMaTEC study group.
Cancer Epidemiol Biomarkers Prev. 2017 Jan;26(1):61-67.
PMID: 27587790

15. **Is sidedness prognostically important across all stages of colorectal cancer?**
Kerr DJ, Domingo E, Kerr R.
Lancet Oncol. 2016 Nov;17(11):1480-1482.
PMID: 27819229

16. **Progressive epigenetic dysregulation in neuroendocrine tumour liver metastases.**
Karpathakis A, Dibra H, Pipinikas C, Feber A, Morris T, Francis J, Oukrif D, Mandair D, Pericleous M, Mohmaduvesh M, Serra S, Ogunbiyi O, Novelli M, Luong T, Asa SL, Kulke M, Toumpanakis C, Meyer T, Caplin M, Beck S, Thirlwell C.
Endocr Relat Cancer. 2017 Feb;24(2):L21-L25.
PMID: 28049633

17. **Somatic POLE proofreading domain mutation, immune response, and prognosis in colorectal cancer: a retrospective, pooled biomarker study.**
Domingo E, Freeman-Mills L, Rayner E, Glaire M, Briggs S, Vermeulen L, Fessler E, Medema JP, Boot A, Morreau H, van Wezel T, Liefers GJ, Lothe RA, Danielsen SA, Sveen A, Nesbakken A, Zlobec I, Lugli A, Koelzer VH, Berger MD, Castellví-Bel S, Muñoz J; Epicolon consortium, de Bruyn M, Nijman HW, Novelli M, Lawson K, Oukrif D, Frangou E, Dutton P, Tejpar S, Delorenzi M, Kerr R, Kerr D, Tomlinson I, Church DN.
Lancet Gastroenterol Hepatol. 2016 Nov;1(3):207-216.
1253(16)30014-0.
PMID: 28404093

DISSEMINATION

The overall strategic goal for DoMore!'s communication has been to increase awareness about AI among both professionals as the general public.

Through active promotion efforts towards journalists and editors in the top Norwegian newsrooms, the unit has reached some of the country's most important media. The results are among others a front-page report and a corresponding TV feature in VG, two front-page reports in Norwegian daily newspaper Dagbladet, a story on the Norwegian public broadcaster NRK's website, and a mentioning on the cover of Aftenposten's weekly A-Magasinet.

RAISING THE LEVEL OF PUBLIC KNOWLEDGE

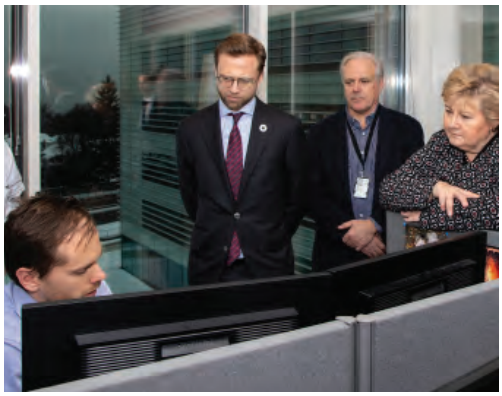
The amount of interest we have seen so far indicates that we have been successful with our strategic communication goals. More than close to 40 presentations and more 50 news reports in national and international media over the past two and a half years have helped to position DoMore! at the public forefront of Norwegian AI-based cancer research.

ICGI's Unit for dissemination and visualization provides the communication strategy. The team of communicators consists of a designer, a 3D designer, a web developer, and a writer and are contributing to the dissemination of DoMore! through various platforms and media channels. In addition to the website, we have so far shared popularizations of our research on the Institute's YouTube channel with close to 45,000 subscribers, on Facebook, Twitter and more recently also on Instagram.

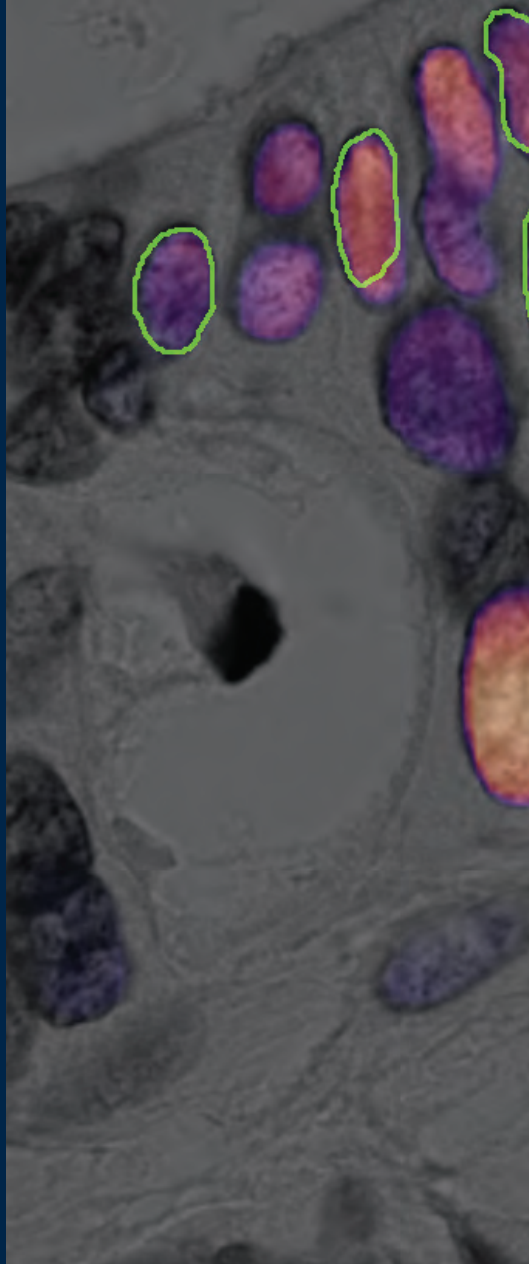
MEDIA MONITORING

To increase the efficiency of the outreach efforts, have we used press release tools such as the publication site Mynewsdesk and the media monitoring service Meltwater.

In the next years, we will continue to emphasize national and international media outreach, in addition to maintaining the activities we have described. Simultaneously, we will continue working towards increasing interest among national and international academic communities. We will also continue the work to establish and level an understanding for DoMore! in our own the hospital organisation, with future implementation in mind.



In the beginning of February 2019 the Norwegian government launched the startup to develop a National strategy for using Artificial Intelligence (AI). The announcement was made during a visit to the Institute for Cancer Genetics and Informatics, where parts of the DoMore! project were presented. The visit included a tour of our laboratories as well as a presentation of how AI can be used in cancer prognostics. The left-most image shows, (from left, at back): Project leader, Håvard E. Greger Danielsen, Cathrine Lofthus, Director South-Eastern Norway Regional Health Authority , Sigbjørn Smeland, Head of the Cancer Clinic, Oslo University hospital, Erna Solberg, Prime Minister, Nikolai Astrup, Digitisation Minister. Bioengineer Maria Isaksen (left image), Post.doc. Manohar Pradhan (center image) and System developer/ Post.doc, Sepp de Raedt demonstrated their work.



DoMore!

- A project for cancer patients and society

The DoMore! project was in 2016 selected as one of the Norwegian Research Council's Lighthouse projects to solve large societal challenges using cutting-edge technology. The DoMore! project will run until 2021, led by the Institute for Cancer Genetics and Informatics (ICGI) at Oslo University Hospital.

www.domore.no

Oslo, February 2019