

Whole Slide Image (WSI) learning and applications

Tristan Lazard

Résumé

Whole Slide Images are numerical version of microscopic slides. They are taken in routine in order to diagnose cancer, either thanks to a biopsy or a piece of the ablated tissue during a surgical operation. Thanks to the high magnification and resolution at which the images are numerised, they are mines of useful information: at the cell scale, at the tissue scale and even at the broader scale of organs. My work during this first year of PhD focused on leveraging this information to predict clinical outcomes and cellular pathways dysfunction. The nature of the WSI itself makes this task quite challenging. This report will be 3-folded : I will begin with reporting peculiarities of WSI and the solutions that have been adopted in the litterature. In a second time I will describe the applications problems that motivated my work as well as the solution I implemented to tackle them. I will finally present the results I obtained as well as the perspectives it gives us.

Chapitre 1

Biological and methodological challenges

Biological motivation : the prediction of the Homologous recombination deficiency (HRD)

Breast cancer is a very heterogeneous disease with four major molecular classes (luminal A, B, HER2 enriched and triple-negative) each of which benefiting from different therapeutic approaches figure. Early breast cancer patients benefit from an overall survival of 70 to 80%. However, the advanced stage with metastatic disease is incurable.

On average, a human cell undergoes 10 double-stranded breaks in its chromosomes. If these breaks occurs in important regions such as for example an anti-oncogene, if they are not repaired, can lead to carcinogenesis. The homologous recombination pathway is one of the three main cellular processes that aim at repairing double stranded lesions of the DNA. It occurs in up to 50% of triple-negative breast cancers, and in up to 18% of breast cancer all molecular classes combined (Turner 2017)

Many proteins take place in the formation of the complex responsible for the double-stranded DNA repair: BRCA1, BRCA2, RAD51C or PALB2 for example. Thus, the paths leading to HRD are multiple, and can be either acquired in a hereditary context or in the tumor during carcinogenesis. The majority of hereditary BRCA1 cancers are triple negative (TN) and 50 to 60% of TN harbor a genomic profile of HRD. HRD also exists in luminal B (Manié et al. 2016) or HER2 (Ferrari et al. 2016) tumors associated with hereditary BRCA2 mutations or in a sporadic context. This defect is therefore difficult as its causes are multiple and even for some cases unknown.

As evoked, HR pathway is not the only double-stranded DNA repair pathway (Rodrigues et al., s. d.), there is also the Non-homologous end joining (NHEJ) and the alternative end-joining (AltEJ). The latter can be targeted by various drugs such as PARPi or platinum salts, in an inhibition purpose. The cells that accumulate the defect of both HR and AltEJ would suffer from too many DNA defects and eventually die, and these cells are supposed to be the cancerous cells ! Therefore, PARPi and platinum salts are

promising drugs in the treatment of HRD tumors, and already have documented proven effects (Tutt et al. 2018).

BRCA1 and BRCA2 mutations are predictive markers for response to PARPi (Paluch-Shimon et Evron 2019) and platinum salt (Tutt et al. 2018). In that context, the screening for BRCA1 and BRCA2 mutations for all luminal B and triple negative breast cancers, that represent 35% and 15 to 18% of all breast cancers, respectively, is a real challenge in clinical practice.

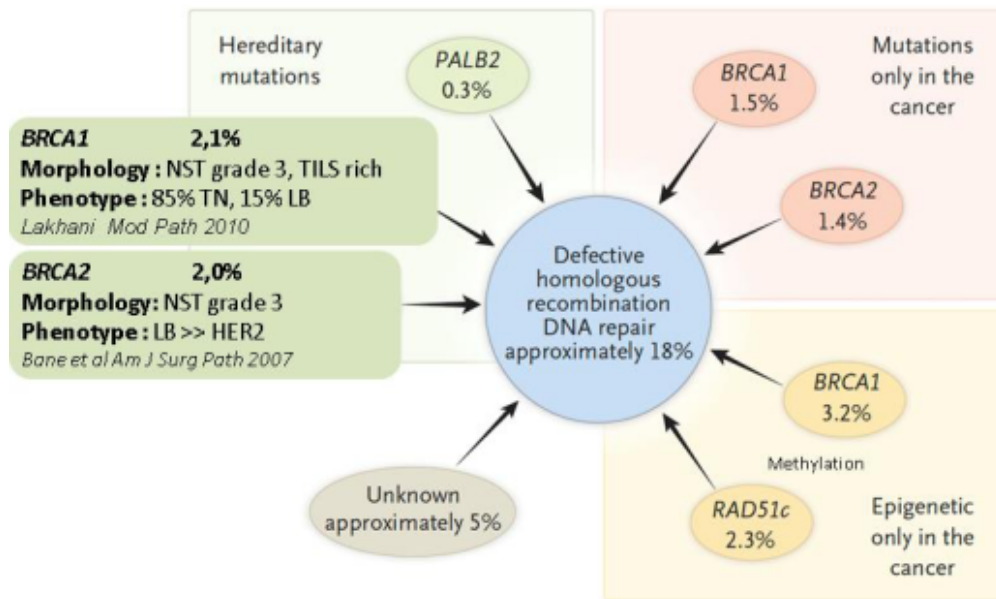


FIGURE 1.1 – of the HRD. Adapted from (Turner 2017). NST: Non special type. IHC: Immunohistochemistry. TILs: Tumor Infiltrating Lymphocytes. LB: Luminal B

Furthermore, no specific morphological patterns or phenotypic characteristics distinguish carcinomas with HRD as a result of either germline or somatic BRCA1/2 mutations in one hand and HRD due to other mechanisms (BRCA1 methylation, RAD51C mutation...) in the other hand. Similar as for lung (Coudray et al. 2018), mesothelioma (Courtiol et al. 2018), gastrointestinal (Kather et al. 2019), (Schmauch et al. 2020) tumors, the development of machine- learning models could predict clinico-molecular characteristics of breast tumors from digitalized slides.

The aim of this study is to predict HR status from digitalized slides of breast tumors (triple negative and luminal B breast cancer) using a machine learning approach in order to identify patients more prone to have an HRD defect and to help to better select patients for indications of PARPi and platinum salts. This is **binary classification** problem.

HRD, because of its diverse possible origins, is a challenge to measure precisely. The most precise way of doing so is by conducting functional genomic analysis on in-vitro cell-lines :

nothing close to a possible clinical application. However some proxy measurement have been created, validated by functional analysis ((Birkbak et al. 2012), (Abkevich et al. 2012)). The measure we use as an approximation of the HR state has been developed by a team of the Curie Institute (Popova et al. 2012). It is based on SNP arrays -the DN being taken from frozen tissue-, and measures the number of Large Scale state Transition (LST) present which are clues of an inappropriate recombination between chromosome segments. The HR binary status is inferred from this measure (threshold at 20 LST): a high LST is associated with HRD whereas a low LST is associated with HRP. This measure is applicable in a clinical context, by preleving DNA from frozen tissue samples, and therefore this is this variable that we aim to predict. In the latter, we will refer to the HRD status as LST-status, but it is important to keep in mind this approximation, that may lead to “noisy” labels.

The Whole slides Images : a non-standard input.

To make or confirm the diagnosis of a cancer, a microscopic analysis of ill tissues and cells is prepared and investigated by expert pathologists. This analysis is performed on a sample of tissue (taken from a biopsy or during a surgical operation) that is stained and mounted on a glass-slide. These slides are prepared in routine in a department of pathology, and usually examine under the microscope. Eventually, it is possible to use a digital slide scanner to numerize the sample. These images are called Whole Slide Images (WSI) and are the object of study in computational pathology. Fortunately, thanks to the rise of computational pathology and recently deep-learning techniques, this last step of scanning is increasingly done in hospitals, giving rise to bigger and more complete diseases datasets.

General characteristics of WSI

A WSI is usually scanned at a magnification 40X or 20X. This magnification leads to huge images: an example of a WSI as well as their typical sizes can be seen on figure 1.2.

This size leads to several constraints that we had to overcome in order to do image analysis and apply machine learning algorithm:

1. The whole image hardly fits in memory.
 - A specific library has to be used to process such objects (`openslide` or `pyvips`).
 - We have to cut a WSI into small images of the same size that are named **tiles**.
A WSI is, on average, composed of **50000 tiles at the highest resolution**.
2. The signal-to-noise ratio is particularly small. As exemplified in figure 1.2 A. : whereas some instances are interesting for pin-pointing a diagnostic (red tile), a lot more are useless and add noise to the overall image (blue tile).
3. The many-steps protocole to create a WSI leads to as many possible ways to incorporate biases in them. Taking care of these biases is fundamental when treating such datasets.

Because of these particularities, we cannot apply candidly the classical image processing algorithm on WSI.

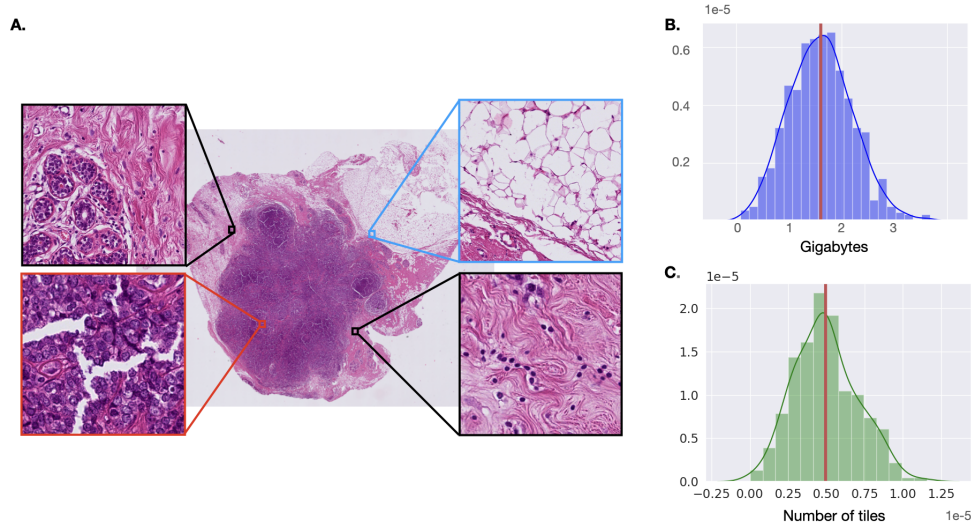


FIGURE 1.2 – **A.** Example of a WSI extracted from our in-house dataset.. **B.** Size of the WSI of this dataset in Gigabytes. **C.** Size of this dataset in number of 256x256 tiles at full resolution.

Slide creation protocole and caveat

As points 1. and 2. are methodocological challenges (in term of algorithm and computation), point 3. is more practical. To understand precisely the bias we could care about and develop an intuition about them, I will here quickly describe how a whole slide is created. As described in (Azevedo Tosta et al. 2019) there is 7 steps in the creation of a slide: **collection**, **fixation**, dehydration and clearing, paraffin embedding, microtomy, **staining** and mounting. The collection describes the way the tissue sample is prelevé (biopsy or surgical prelevement). The fixation aims at stopping biochemical reactions in the tissue and therefore preserve it. dehydration and clearing aim at drying the sample for it to be cut easily, the paraffin embedding as well. The sample block is then cut into thin slides, this is the microtomy (a precise machine is used to do that). Finally the sample is stained with particular dyes, that are supposed to highlight specific tissue components. The samples are eventually mounted on glass slides.

The **bolded** steps are

WSI processing and learning is actually an active research field, and I will try to summarize the different methods developed in the litterature.

WSI processing: a litterature review

As mentioned previously, the most natural way of processing a WSI is to cut it into small, more bearable pieces: the tiles. In my implementations, a tile is an image of 256x256

pixels. By doing so, if we do not consider the spatial distances between tile, a WSI can be represented by the set of its composing tiles. The machine learning task -which is to classify a WSI between HRD and HRP- therefore shrinks to classifying a set of tiles. This problem of classifying sets of instances has been known for a while by the machine learning community as the **multiple-instance learning** problem. It provides a theoretical and practical framework for solving the WSI classification problem. In the following, I will introduce the multiple-instance learning problem, then present how its development has been used and adapted to the WSI classification task. In a third time, I will present the limitations of the methods drawn from this framework and present a few methods that do not fit within it.

The multiple instance learning (MIL) problem.

The multiple instance learning problem is a supervised machine-learning problem with a **weak supervision**. The weak supervision regime can describe different supervision regimes, as the review of Zhou summarizes it (Zhou 2018):

1. Incomplete supervision: only part of the available data is labeled, and this part is usually not big enough to train a good learner.
2. Inexact supervision: the given supervision is concerning only a small part of the input.
3. Inaccurate supervision: the given supervision is noisy. It may suffer from errors.

The standard MIL problem aims at classifying a *bag of instances* or *bag of feature vectors* instead of a single *feature vector*. The supervision associates one bag with one label and we know that only some of the instances contained in the bag are responsible for the label (making the other instances *wrongly* associated with the label). It therefore lays within the second supervision regime that have been mentioned. However WSI classification problems usually have a supervision information sharing the characteristics of all the three points.¹

This setting is fitted to many tasks such as for example drug activity classification. Let’s say for instance that we want to predict the effect of a molecule from its different spatial conformations, it is natural to represent one drug as the bag of its possible spatial conformations. Consequently, the purpose of multiple instance learning is to find which instances are important for the classification task, and aggregate them appropriately.

The standard MIL assumption

In the case of the prediction of the effect of a molecule, if one of the conformation of a molecule has an effect, then the molecule itself has an effect. Therefore, if we make the assumption that each instance have a (hidden) label, then the relationship between the bag label and the instances labels is straightforward! If we set as 1 the label “has an effect”, the label of the whole bag is exactly the maximum of the instances labels.

1. **1.** Labels are very costly to obtain -LST signature for example- whereas images that are taken in routine, are available in substantial quantities. **3.** In many problems, the annotation is done by experts and suffers a part of subjectivity.

This assumption is called the **standard MIL assumption**. As done in the review work of Foulds (Foulds et Frank 2010), we formalize the problem (and set here the notation for the following of the report).

We set $X = \{x_1, \dots, x_n\}$ to be a *bag* object containing $x_i \in \mathcal{X}$ instances. \mathcal{X} is the feature space in which instances are embedded, usually $\mathcal{X} = \mathbf{R}^f$. The standard MIL assumption has been stated to solve a binary classification; therefore let $\Omega = \{+, -\}$ be the label space. As said previously we hypothesize that it exists a (unknown) function $g : \mathcal{X} \rightarrow \Omega$ that maps each instance to its hidden label. We are looking for a function $\nu : \mathbb{N}^{\mathcal{X}} \rightarrow \Omega$, called the *concept*, that will map a bag of instances to its label. The concept specify the relationship that exists between the instances labels and the bag label.

Following the standard MIL assumption, and given the \vee is the OR operator, we have the equivalency relationship :

$$\nu(X) \Leftrightarrow (g(x_1) \vee g(x_2) \vee \dots \vee g(x_n))$$

This assumption is really well suited to the molecule activity prediction task. It may also be adapted for some of the Whole Slide Images tasks. For instance, if the classification problem is to classify a WSI as cancerous, then it is common sense to classify the WSI as cancerous if we can find inside it a patch containing cancerous cells.

However the concepts that can be leveraged by this assumption may not be rich enough to solve other tasks. Just consider for example the classification problem of grading a tumor. One of its key features is the mitotic index, that is, the count of cells in a division state, inside the tumor. One tile containing a mitose won't be enough to classify a WSI as high-grade ! In these case we could need a more broader concept ν , for example a concept based on a count of instances :

$$\nu(X) \Leftrightarrow \sum \mathbf{1}_{\{g(x_i)=+\}} > s$$

Where $+$ is the label associated to an instance containing a mitotic figure, and s a threshold fixing the shift between high and low grade tumors. Foulds et al. (Foulds et Frank 2010) review the possible MIL assumptions, that shaped the different solutions of WSI classification.

From standard MIL to WSI-MIL

If we want to predict wether or not a tissue slide is cancerous, then the standard MIL assumption is the exact hypothesis to do. We can draw from this a general framework to solve the WSI classification problem. If we follow the logic of the standard MIL assumption, in order to output a WSI prediction, we will need three functions:

1. One function \mathcal{F} that will classify the *instances*.
2. One function \mathcal{P} that will aggregates the instances predictions, the *concept*.
3. Finally a function \mathcal{C} that will generate a prediction from the concept, and therefore classify the whole bag.

The compositions of the three functions creates a concept and classify it.

$$\mathcal{C} \circ \mathcal{P} \circ \mathcal{F}(X) = \hat{y} \in [0, 1]^C$$

with C being the number of classes. If $\mathbf{2}$. is the max function and $\mathbf{3}$. is the identity, the concept is the one of the standard assumption. We can see that if \mathcal{C} , \mathcal{P} and \mathcal{F} are derivable, then we can compute a classical classification loss $L(X) = L(y, \mathcal{F}(X)) = L(y, \hat{y})$, y being the label of X , and optimize the parameters of each of this function using the Stochastic Gradient Descent (SGD) algorithm.

In the case of a *more general* WSI classification problem, we would prefer to use a *more general* assumption than the standard one (or even the “counting” assumption cited above).

The generalization of the standard MIL assumption provides a framework for innovation and most of the algorithm that have been developed for WSI classification aim at improving one of the three presented functions, or several ones independantly.

The work by Campanella et al. (Campanella et al. 2019), that is currently the work applying a MIL algorithm to the largest WSI database, use a slightly modified version of the standard MIL algorithm. They set \mathcal{F} to a ResNet (a standard convolutional neural network architecture) that outputs a score (which can be interpreted as a classification score). The concept $\mathcal{P} = \operatorname{argmax}_{x_i} \mathcal{F}(x_i)$ picks the best scoring tile of the WSI, and they finally generate the slide-level prediction score with an other ResNet (\mathcal{C}). In a second step, they freeze \mathcal{F} , set $\mathcal{P} = \operatorname{top-k}_{x_i}(\mathcal{F}(x_i))$ and learn \mathcal{C} as a recurrent neural network, that aggregates the top-k tiles according to the \mathcal{F} scoring function. This last step is reported to add robustness to the results.

The work by Courtiol et al. (Courtial et al. 2018) develops a MIL algorithm by changing mostly the \mathcal{P} concept function, that is the pooling function. \mathcal{F} is here again a ResNet. This time, as Campanella and team did, they choose $\mathcal{P} = (\operatorname{top-k}_{x_i} \mathcal{F}(x_i), \operatorname{low-k}_{x_i} \mathcal{F}(x_i)) \in \mathbf{R}^{2k}$ as the pooling function. \mathcal{C} is parametrized as a Multilayer perceptron. Courtial and his team have been inspired by the work from Durand et al. (Durand et al. 2017), where they show that using *negative* instances in the final prediction is beneficial to the overall classification. This seems pertinent in natural images classification (if in a maintain scene a girafe is present, it should be normal to have the instance **girafe** decreasing the bag probability of being a mountain scene). however often, depending on the way WSI are tiled, the worst scored tiles are artifacts usually associated with the preparation of the sample (badly focused or inked regions for example).

The proposition of Ilse et al. (Ilse, Tomczak, et Welling 2018) is by far the most popular one and is in many papers the baseline to compete with. This work focus on the pooling function: on the contrary to the two previous method, Ilse proposes in this work to *learn* this pooling function. \mathcal{F} is a parametrized by a neural network that outputs not one score, but a low dimensional embedding of the input. Then the pooling function \mathcal{P} is a weighted sum of these low dimensional embeddings, the weights being the output of a neural networks (parametrized by $\mathbf{V} \in \mathbb{R}^{L \times M}$ and $\mathbf{W} \in \mathbb{R}^{L \times 1}$).

$$\mathcal{P}(\mathcal{F}(X)) = \sum_{i=1}^N a_i \mathcal{F}(x_i)$$

with

$$a_i = \frac{\exp \{ \mathbf{W}^\top \tanh(\mathbf{V} \mathbf{h}_i^\top) \}}{\sum_{j=1}^N \exp \{ \mathbf{W}^\top \tanh(\mathbf{V} \mathbf{h}_j^\top) \}}$$

This brings flexibility to the framework: the pooling function can then be closer to the **max** function or closer to the **mean** function depending on the problem. It is suited (theoretically) to solve a broader range of MIL assumptions.

All of these algorithm share also the same interest of being easily **interpretable**. In fact, in each of them, tiles participates in the bag representation in proportion to the score that is output, either by \mathcal{F} or \mathcal{P} . These scores may therefore serve as a key for interpreting the prediction, by extracting the most important tiles or creating heatmaps of “importantness”.

Finally, many other work produced variations of the presented algorithm, mainly by adapting the pooling function to the task or data available (Li et Eliceiri 2020), (Gildenblat et al. 2020), (Bhattacharjee et al. 2020).

To conclude, multiple instance learning provides a good rational to solve the WSI classification problem. However it seems like the MIL-based solution have been adopted mostly because of the resemblance between a tiled WSI and the bag of instance used in MIL. On the contrary to the set of molecule conformations described earlier, the set of tiles that compose a WSI are spatially related. In the previous implementations, the image nature of the WSI is not taken into account; more than that, the MIL pooling function have been specifically developped to be permutation invariant ((Zaheer et al. 2018), (Sannai, Takai, et Cordonnier 2019)), meaning that the spatial relationship between instances is purposely disregarded in these models.

Fewer works are focusing on treating the WSIs not as a bag of instances but as an image. In the article of Pinckaers et al. (Pinckaers, Ginneken, et Litjens 2019), succeed in applying a classical CNN on very large portions of WSI. By streaming these very large images (8092x8092 pixels) through the CNN and reconstructing the activation maps during the stream, they have been able to achieve very good results on the *CAMELYON16* benchmark dataset. This engineering *tour de force* allowed them to process images that would require GPU units with 825Gigabytes of memory (per batch) in the classical framework, without significant loss in the performances. Saliency maps can moreover be extracted to help the interpretation of the classification results.

An other work by Tellez et al. ((Tellez et al. 2019), (Tellez et al. 2020)) took the side of applying a convolution after the *compression* of the WSI. As in the previous works, a WSI is first tiled. These tiles are used to train an unsupervised representation learning algorithm (this paper providing, beside its compression idea, a comparative study of three unsupervised learning algorithm in the histopathology context). The encodings extracted from them, of a lower dimension than the initial tiles, are then stacked according to their

place in the initial WSI. Finally a CNN is used on top of this compressed version of the WSI for classification purposes.

While these two last proposition are adresssing the lack of contextual information of the various MIL algorithms, they still share some drawbacks. The implementation of the first solution is not part of the usual deep-learning libraries, consequently making further development and innovation hard and less flexible. Both of them need the input WSI to be cropped to a standard size, and therefore cannot appropriately account for the high diversity in high and shape of the slides. Finally, we expect the lack of spatial context of MIL models to bring somehow a plus for training with a small dataset. In fact, the shape and surface area of the tissue in the slides are a great variability factor that, we think, would easily cause overfitting.

In the next part of this report I will present the model I used and improved, as well as the results I obtained on several classification problems.

Dont forget to evoke when presenting my model : - the problem of Attention DEep mil for interpreting the scores (if the instances representation do not share the same magnitude)
- The way I changed the model in order to improve interpretation (peut etre qu'une normalisation futée à la fin de la représentation de basse dimension serait suffisant pour régler le pb ?) - Develop the instance-dropout part (re-launch experiments on big dataset of challenge) - creates T-sne representation of each head of attention

Biological Motivations

Describe here the HRD-BRCAness problem.

The developed solution:

Features extraction

Multiple-instance learning model

Results

HRD

standard variables (grade etc...)

Challenge

Perspectives

Chapitre 2

Introduction to WSI classifications

- Abkevich, V., K. M. Timms, B. T. Hennessy, J. Potter, M. S. Carey, L. A. Meyer, K. Smith-McCune, et al. 2012. « Patterns of Genomic Loss of Heterozygosity Predict Homologous Recombination Repair Defects in Epithelial Ovarian Cancer ». *British Journal of Cancer* 107 (10): 1776-82. <https://doi.org/10.1038/bjc.2012.451>.
- Azevedo Tosta, Thaína A., Paulo Rogério de Faria, Leandro Alves Neves, et Marcelo Zanchetta do Nascimento. 2019. « Computational Normalization of H&E-Stained Histological Images: Progress, Challenges and Future Potential ». *Artificial Intelligence in Medicine* 95 (avril): 118-32. <https://doi.org/10.1016/j.artmed.2018.10.004>.
- Bhattacharjee, Kamanasish, Arti Tiwari, Millie Pant, et Chang Wook Ahn. 2020. « A Pooling Function Based on Differential Evolution for Multiple Instance Learning », 6.
- Birkbak, Nicolai J., Zhigang C. Wang, Ji-Young Kim, Aron C. Eklund, Qiyuan Li, Ruiyang Tian, Christian Bowman-Colin, et al. 2012. « Telomeric allelic imbalance indicates defective DNA repair and sensitivity to DNA-damaging agents ». *Cancer Discovery* 2 (4): 366-75. <https://doi.org/10.1158/2159-8290.CD-11-0206>.
- Campanella, Gabriele, Matthew G. Hanna, Luke Geneslaw, Allen Mirafior, Vitor Werneck Krauss Silva, Klaus J. Busam, Edi Brogi, Victor E. Reuter, David S. Klimstra, et Thomas J. Fuchs. 2019. « Clinical-Grade Computational Pathology Using Weakly Supervised Deep Learning on Whole Slide Images ». *Nature Medicine* 25 (8): 1301-9. <https://doi.org/10.1038/s41591-019-0508-1>.
- Coudray, Nicolas, Paolo Santiago Ocampo, Theodore Sakellaropoulos, Navneet Narula, Matija Snuderl, David Fenyő, Andre L. Moreira, Narges Razavian, et Aristotelis Tsi-rigos. 2018. « Classification and Mutation Prediction from Non-Small Cell Lung Cancer Histopathology Images Using Deep Learning ». *Nature Medicine* 24 (10): 1559-67. <https://doi.org/10.1038/s41591-018-0177-5>.
- Courtiol, Pierre, Eric W. Tramel, Marc Sanselme, et Gilles Wainrib. 2018. « Classification and Disease Localization in Histopathology Using Only Global Labels: A Weakly-Supervised Approach ». *arXiv:1802.02212 [cs, stat]*, février. <http://arxiv.org/abs/1802.02212>.
- Durand, Thibaut, Taylor Mordan, Nicolas Thome, et Matthieu Cord. 2017. « WILDCAT: Weakly Supervised Learning of Deep ConvNets for Image Classification, Pointwise Localization and Segmentation ». In, 642-51. http://openaccess.thecvf.com/content_cvpr_2017/html/Durand_WILDCAT_Weakly_Supervised_CVPR_2017_paper.html.
- Ferrari, Anthony, Anne Vincent-Salomon, Xavier Pivot, Anne-Sophie Sertier, Emilie Thomas, Laurie Tonon, Sandrine Boyault, et al. 2016. « A Whole-Genome Sequence and

- Transcriptome Perspective on HER2-Positive Breast Cancers ». *Nature Communications* 7 (1): 12222. <https://doi.org/10.1038/ncomms12222>.
- Foulds, James, et Eibe Frank. 2010. « A Review of Multi-Instance Learning Assumptions ». *The Knowledge Engineering Review* 25 (1): 1-25. <https://doi.org/10.1017/S026988890999035X>.
- Gildenblat, Jacob, Ido Ben-Shaul, Zvi Lapp, et Eldad Klaiman. 2020. « Certainty Pooling for Multiple Instance Learning ». *arXiv:2008.10548 [Cs]*, août. <http://arxiv.org/abs/2008.10548>.
- Ilse, Maximilian, Jakub M. Tomczak, et Max Welling. 2018. « Attention-based Deep Multiple Instance Learning ». *arXiv:1802.04712 [cs, stat]*, juin. <http://arxiv.org/abs/1802.04712>.
- Kather, Jakob Nikolas, Alexander T. Pearson, Niels Halama, Dirk Jäger, Jeremias Krause, Sven H. Loosen, Alexander Marx, et al. 2019. « Deep Learning Can Predict Microsatellite Instability Directly from Histology in Gastrointestinal Cancer ». *Nature Medicine* 25 (7): 1054-6. <https://doi.org/10.1038/s41591-019-0462-y>.
- Li, Bin, et Kevin W. Eliceiri. 2020. « Dual-stream Maximum Self-attention Multi-instance Learning ». *arXiv:2006.05538 [cs]*, juin. <http://arxiv.org/abs/2006.05538>.
- Manié, Elodie, Tatiana Popova, Aude Battistella, Julien Tarabeux, Virginie Caux-Moncoutier, Lisa Golmard, Nicholas K. Smith, et al. 2016. « Genomic Hallmarks of Homologous Recombination Deficiency in Invasive Breast Carcinomas ». *International Journal of Cancer* 138 (4): 891-900. <https://doi.org/10.1002/ijc.29829>.
- Paluch-Shimon, Shani, et Ella Evron. 2019. « Targeting DNA Repair in Breast Cancer ». *The Breast* 47 (octobre): 33-42. <https://doi.org/10.1016/j.breast.2019.06.007>.
- Pinckaers, Hans, Bram van Ginneken, et Geert Litjens. 2019. « Streaming convolutional neural networks for end-to-end learning with multi-megapixel images ». *arXiv:1911.04432 [cs]*, novembre. <http://arxiv.org/abs/1911.04432>.
- Popova, T., E. Manie, G. Rieunier, V. Caux-Moncoutier, C. Tirapo, T. Dubois, O. Delattre, et al. 2012. « Ploidy and Large-Scale Genomic Instability Consistently Identify Basal-Like Breast Carcinomas with BRCA1/2 Inactivation ». *Cancer Research* 72 (21): 5454-62. <https://doi.org/10.1158/0008-5472.CAN-12-1470>.
- Rodrigues, M, É Manié, P Cottu, V Diéras, G Bataillon, T Popova, et M H Stern. s. d. « Ciblage des défauts de la recombinaison homologue par les inhibiteurs de PARP ». *MISE AU POINT*, 6.
- Sannai, Akiyoshi, Yuuki Takai, et Matthieu Cordonnier. 2019. « Universal approximations of permutation invariant/equivariant functions by deep neural networks ». *arXiv:1903.01939 [cs, stat]*, septembre. <http://arxiv.org/abs/1903.01939>.
- Schmauch, Benoît, Alberto Romagnoni, Elodie Pronier, Charlie Saillard, Pascale Maillé, Julien Calderaro, Aurélie Kamoun, et al. 2020. « A Deep Learning Model to Predict RNA-Seq Expression of Tumours from Whole Slide Images ». *Nature Communications* 11 (1): 3877. <https://doi.org/10.1038/s41467-020-17678-4>.

- Tellez, David, Diederik Hoppener, Cornelis Verhoef, Dirk Grunhagen, Pieter Nierop, Michal Drozdal, Jeroen van der Laak, et Francesco Ciompi. 2020. « Extending Unsupervised Neural Image Compression with Supervised Multitask Learning ». *arXiv:2004.07041 [Cs, Eess]*, avril. <http://arxiv.org/abs/2004.07041>.
- Tellez, David, Geert Litjens, Jeroen van der Laak, et Francesco Ciompi. 2019. « Neural Image Compression for Gigapixel Histopathology Image Analysis ». *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1-1. <https://doi.org/10.1109/TPAMI.2019.2936841>.
- Turner, Nicholas C. 2017. « Signatures of DNA-Repair Deficiencies in Breast Cancer ». *The New England Journal of Medicine* 377 (25): 2490-2. <https://doi.org/10.1056/NEJMcibr1710161>.
- Tutt, Andrew, Holly Tovey, Maggie Chon U. Cheang, Sarah Kernaghan, Lucy Kilburn, Patrycja Gazinska, Julie Owen, et al. 2018. « Carboplatin in BRCA1/2-mutated and triple-negative breast cancer BRCAness subgroups: the TNT Trial ». *Nature Medicine* 24 (5): 628-37. <https://doi.org/10.1038/s41591-018-0009-7>.
- Zaheer, Manzil, Satwik Kottur, Siamak Ravanbakhsh, Barnabas Poczos, Ruslan Salakhutdinov, et Alexander Smola. 2018. « Deep Sets ». *arXiv:1703.06114 [cs, stat]*, avril. <http://arxiv.org/abs/1703.06114>.
- Zhou, Zhi-Hua. 2018. « A Brief Introduction to Weakly Supervised Learning ». *National Science Review* 5 (1): 44-53. <https://doi.org/10.1093/nsr/nwx106>.