

# A web-based framework for visualization, annotation, and automatic exploitation of whole-slide images

Arnaud Babilone<sup>1</sup>, Samuel Cop<sup>1</sup>, Rafik Rassaa<sup>1</sup>, Natacha Rocks<sup>2</sup>, Loïc Rollus<sup>1</sup>, Didier Cataldo<sup>2</sup>, Raphaël Marée<sup>3</sup>



<sup>1</sup>MSc. student at Dept. EE & CS, University of Liège, Belgium

<sup>2</sup>Laboratory of Tumor Development Biology, GIGA-R, University of Liège, Belgium

<sup>3</sup>GIGA Bioinformatics Core Facility, GIGA, University of Liège, Belgium

## Introduction

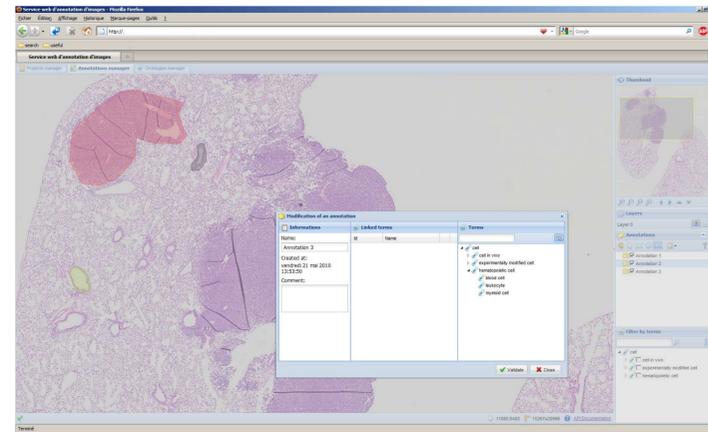
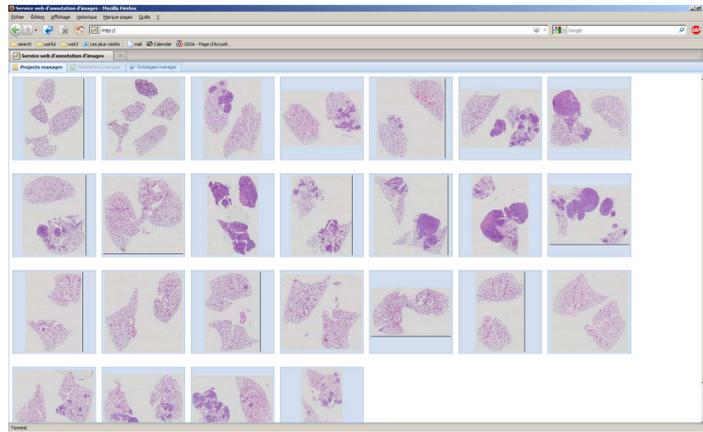
With the recent advent of biosensors and digital image acquisition technologies, scientists are now able to routinely generate gigabytes or even terabytes of imaging data. However, human interpretation of such large-scale datasets is very time-consuming and somehow subjective. This stresses the need for replicable and generic computational methods to automate the extraction of useful, quantitative, information from these images.

Towards that goal, we have identified a strong need for intuitive and efficient computerized tools to foster collaboration between biomedical researchers working with virtual microscopy and computer scientists developing machine learning and computer vision methods.

In this context, we present the ongoing development of a web-based software platform for remote annotation, visualization, and automated analysis of whole-slide images.

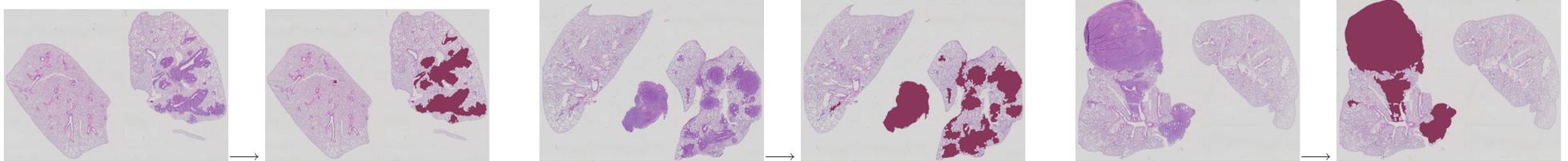
## Whole-slide Visualization and Annotation On the Web

Our lightweight whole-slide image server and viewer uses Java/JavaScript technologies and a relational database to allow collaborative annotations of images using ontology terms.



## Tree-based Learning for Automatic Segmentation and Quantification of Tissues

Our recent generic machine learning algorithm using random subwindows and tree-based approaches [1,2] exploits manual annotations to learn a segmentation model based on local raw pixel values. The model can be later applied on new images for automatic tissue quantification. It will be integrated into the web platform in the near future.



## Distributed and Incremental Content-Based Image Retrieval

The potential number of whole-slide images, their sizes, their availability across different centers or hospitals, and frequent additions (as new tissues are scanned) makes efficient, distributed and incremental CBIR a useful tool to help researchers and pathologists to explore their data and support their findings or diagnosis. We illustrate the potential of our recent CBIR approach [3] to index roughly about 53000  $256 \times 256$  pixels tiles extracted from 8 whole-slide images of experimental lungs generated for a cancer project.



## References

- 1 Dumont et al., *Fast Multi-Class Image Annotation with Random Subwindows and Multiple Output Randomized Trees*, VISAPP 2009.
- 2 Marée et al., *Random Subwindows and Extremely Randomized Trees for Image Classification in Cell Biology*, BMC Cell Biology 2007.
- 3 Marée et al., *Incremental Indexing and Distributed Image Search using Shared Randomized Vocabularies*, MIR 2010.

## Acknowledgments

Raphaël Marée is supported by the GIGA (University of Liège) with the help of the Walloon Region and the European Regional Development Fund. This poster presents research results of the Belgian Network BIOMAGNET (Bioinformatics and Modeling: from Genomes to Networks), funded by the Interuniversity Attraction Poles Programme, initiated by the Belgian State, Science Policy Office. This work is supported by the European Network of Excellence, PASCAL2.

