Master internship @ Paris Brain Institute, France
Inria team – "Aramis lab"

Crohn-AI

Unsupervised Omics data clustering as explainable Artificial Intelligence to identify Crohn Disease patients with high risk of severe disease and treatment failure.

Contacts: Daniel Racoceanu - daniel.racoceanu@sorbonne-universite.fr
Mehdi Ounissi – mehdi.ounissi@icm-institute.org
Hosting institute: Paris Brain Institute (Institut du Cerveau - ICM), ARAMIS team, Hôpital Pitié Salpêtrière, 47 Bd. Hôpital, 75013 Paris.
Keywords: Crohn's disease, computational histopathology, whole slide image analysis, molecular parameters, deep learning, explainable artificial intelligence, risk factors, big data analytics.

Project background: Crohn's disease (CD) is an inflammatory bowel disease with an estimated life-long incidence rate close to 2/1.000 in developed countries (1). About 120.000 people are affected in France and several millions around the world. CD most often progresses by relapses interspersed with remissions. Chronic inflammation leads with time to irreversible destruction of the digestive tract with stenosis and fistulas, resulting in loss of function and the need for surgery (2). Remissions and relapses can be numerous or on the contrary rare depending on the patients with resulting various gut damage scores (3). In addition, CD can be associated with extra-digestive symptoms, undernutrition, growth failure in children and gut cancer in adults. Finally, the response to various medical treatments varies greatly from one person to another (2).

Several clinical factors are statistically associated with worse outcome (2). Age (under 40 years), perineal involvement, cigarette smoking, steroid use at the first attack, ileal and upper digestive involvement together with low education level are associated with poor prognosis. Unfortunately, these factors are poor predictors. Researchers thus turned to serological (5) and molecular signatures to predict the disease course (for review see ref 4, 6). Despite some successes, genetic polymorphisms are insufficient to provide a manageable classification of patients (7-9). Transcriptomic information has been exploited to predict clinical outcomes (10-12) or response to treatment (13). Microbiome has also been used alone or with transcriptome (10, 14). Finally, a methylome study is also available (15). At the end, the AUC values of the prognostic tests available today range from 0.69 to 0.81. Such values are usually considered as not adequate for personalized medicine at the individual level.

Disease course thus remains - in large part – unpredictable, making people undertreated with the risk of gut destruction or conversely overtreated with the risk of drug side-effects. In both cases, the cost of the disease is inflated. Up to date, histological information has not been exploited in prognostic studies for CD. Nowadays, recent development of Machine Learning (ML) and, in particular, Deep Learning (DL) algorithms, is revolutionizing Whole Slide Image (WSI) analysis and Computational Pathology (CP).

Aim of the project: The objective of this project is to develop an algorithm able to identify CD patients with the highest risk of severe disease and treatment failure based on integrated datasets with molecular parameters recorded at the time of diagnosis (omics data).

This study is based on:

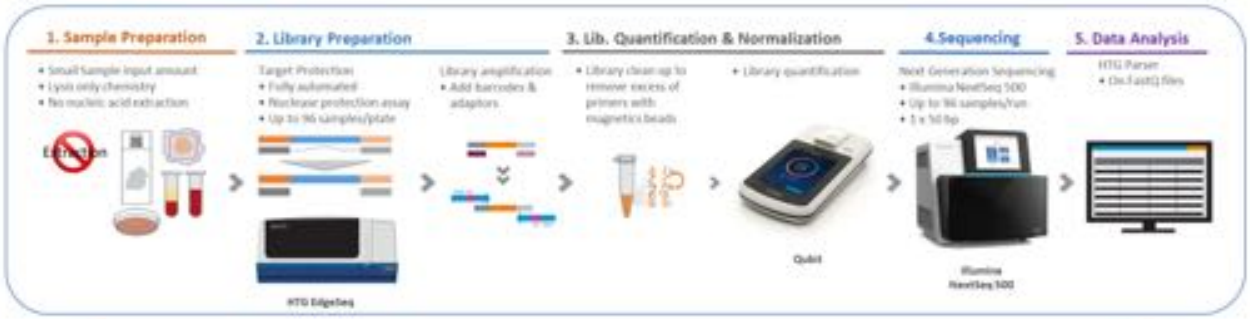
- A retrospective cohort of CD pediatric cases followed in the same hospital and with records of clinical, biological, endoscopic and radiologic data from the initial diagnosis to at least one year of follow-up.
- A newly developed transcriptomic method able to analyze a large panel of genes with a limited quantity of intestinal tissue and with limited biases (the Biopred panel).

Originality and expected impact of the project: add a renewed histological information to the clinical and molecular information. The development of new biomarkers able to predict the disease course and response to treatments would be a key advance in the management of CD.

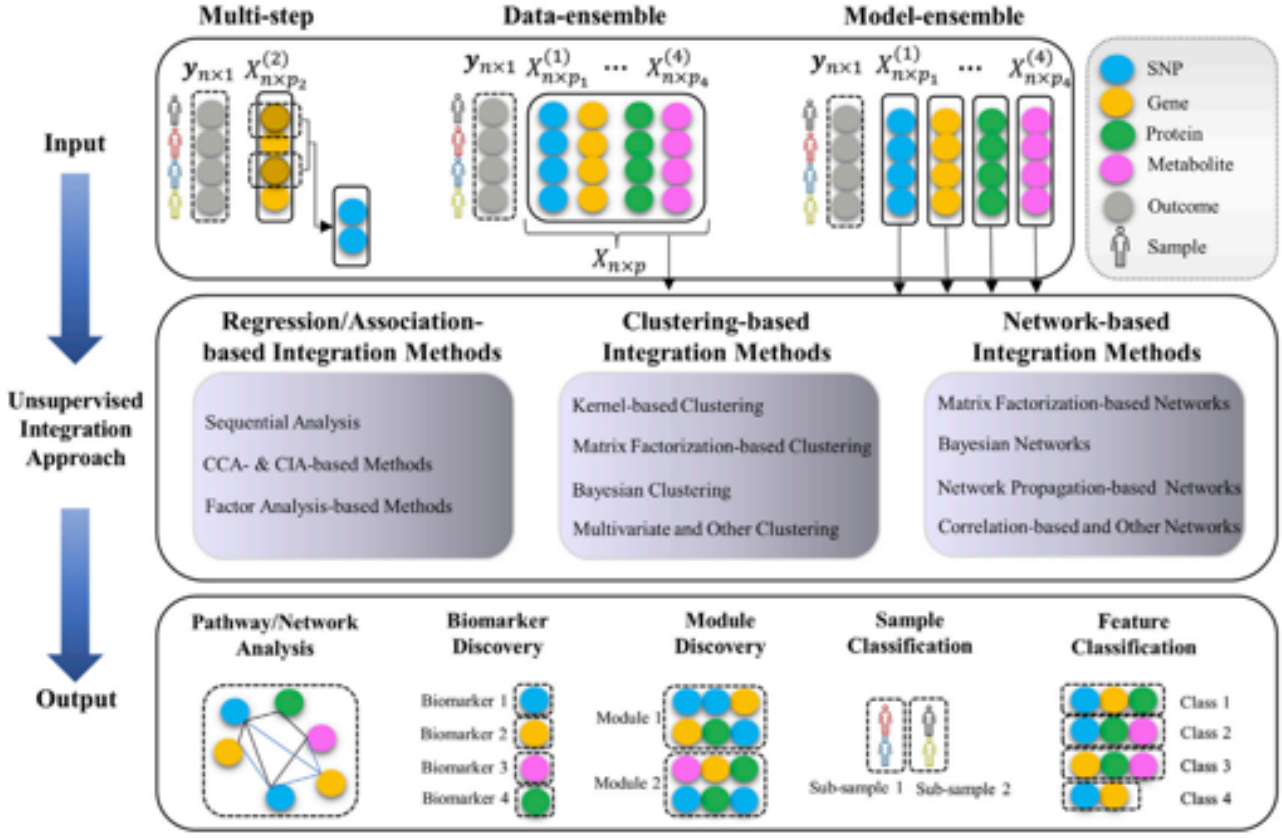
Working plan:

Patients: **436 patients** under the age of 18 years at diagnosis are the studied population. The inclusion criteria are i) a diagnosis of CD made according to ESPGHAN criteria (17) at Robert Debré Hospital and ii) a first biopsy available and iii) at least one year of follow up in this hospital. This population is extremely useful for this work for the two following reasons. 1) The available biopsies were performed at the time of diagnosis, most often before any treatment. They reflect data not modified by the various therapies. Such an approach is usually impossible for cohorts from adult hospitals where access to inaugural biopsies is rare. 2) Cumulative clinical data are available, because children and adolescents are usually followed at Robert Debré's hospital until adulthood.

Transcriptome analysis: The BIOPRED panel developed by our partner, Firalis, is a targeted sequencing gene panel of 2155 mRNA identified especially for those diseases in which biologics are indicated, including CD. These mRNAs are associated with inflammatory and pro-inflammatory pathways, immune response pathways, interleukins, chemokines, growth factors, metalloproteinases, and others. Thanks to the HTG EdgeSeq platform, allowing mRNA measurements without RNA extraction step, the BIOPRED assay allows profiling mRNA targets directly in the samples (see figure below). The RNA extraction free test will be performed directly on tissue sections with as little as 10mm2 slide of 5 µm section. We have checked that it is feasible on the available biological material that will be used for the study. The panel has been previously validated on Rheumatoid Arthritis to predict treatment response (18).



The omics data produced must then be categorized. This can be achieved using unsupervised tools based on deep learning(35), see figure below.



Organization of the work and expected results:

Approval by the French regulatory ethic committee has been obtained. Clinical, biological, radiological and endoscopic findings at diagnosis and during follow-up will be recorded on a e-CRF (epidata). In parallel, the biopsies made during the initial colonoscopy will be retrieved, cut, stained (HES) and scanned to be exploited by Racoceanu's group @ Aramis lab. A small part of each biopsy will be sent to Firalis for transcriptomic analyses. Finally, all kinds of information will be treated altogether in order to look for predictive models. The main expected classification outcomes are: i) the development of stenosis and fistulas requiring surgery with time; ii) extra-intestinal symptoms including growth failure; iii) lack of response to the main treatments (exclusive enteral nutrition, steroids, immunosuppressors and biologics).

Suited candidate profile: candidate should have a solid background in Computer Science with particular focus on Machine and Deep Learning, BioMedical (but not only) Image analysis and Mathematical modeling. Important hands-on skills in computer science and data analysis are requested for this ambitious collaborative project. The open-source programming language will be Python, with the major libraries used in Deep learning (PyTorch) and data analysis (Pandas).

The candidate should have a strong interest in data analysis, numerical implementation and dealing with large databases. She/he must also possess good relational and

communication skills with scientists from various disciplines and very good skills in written and spoken English. We are seeking a motivated candidate, with good track record and writing abilities, interested in working on high impact publications. Previous experience in medical and medical data (incl. images) analytics research would also be valued. We encourage application from people with diverse backgrounds, including people with disability.

Partners of the study:

***Hosting team**: Aramis lab - INRIA team, Paris Brain Institute (Inserm, CNRS, APHP, Sorbonne University) (https://institutducerveau-icm.org/en/)*

- Prof. Daniel Racoceanu's group / ARAMIS lab (INRIA team), is a joint team between Paris Brain Institute and INRIA, developing statistical and computational methods for the analysis of whole slide images as designing, building and testing digital models through data analytics. The team aims at building digital models from multimodal databases from patients. The main approaches used by our team are: i) advance ML, including DL algorithms, with algorithms concerning responsible, explainable, interpretable and auditable artificial intelligence (AI); ii) geometric and statistical modeling, and iii) complex network theory (including sparse-sets mathematical morphology). Daniel Racoceanu's group pioneers Computational Pathology (CP), by launching the world first challenges / benchmarks in this area, since 2012[1,2], by a series of high-impact publications (ref 19, 19-25, 34) dedicated to the proposed hybrid approaches (discriminative ML/DL combined with generative approaches). Major projects (FUI FlexMi[3], ANR TecSan MICO[4]) have been led on CP approaches.

***Partner team**: Dpt. pediatric gastroenterology and nutrition and dpt. of pathology, Hôpital Robert Debré, Assistance Publique Hôpitaux de Paris (http://robertdebre.aphp.fr).*

- The principal investigator of this translational clinical research project is Dr Christine Martinez-Vinson, MD, assistant professor in the department of pediatric gastroenterology. Dr Martinez-Vinson is the supervisor of the CD cohort in the department since more than 15 years. She is specialist of biotherapies in CD and as a result, she is expert in the field of clinical trials and medical research (26-29). She will take in charge the clinical research assistant and e-CRF completion together with the supervision of the whole study.
- Christine Martinez-Vinson is assisted by Professors Jean-Pierre Hugot (MD, PhD) and Dominique Berrebi (MD, PhD). JPH is expert in pediatric IBD. He is head of a fundamental research unit at Research on Inflammation Centre devoted to IBD (http://cri1149.fr). He is experienced in coordination of projects at the national and European levels including AI project (30). DB is the head of the department of pathology at Robert Debré Hospital. She will be in charge of biopsies management and will contribute to the interface. She has published in the field of CD, with Dr Martinez-Vinson and Prof Hugot (see ref 31-33).

To apply or for further information, please send an email with CV, cover letter to daniel.racoceanu@sorbonne-universite.fr.

---

[1] *MITOS 2012 (mitosis detect.) – 1st DP challenge, ICPR 2012, Tsukuba, Japan (130 particip., lab & institutes)*
[2] *Atypia assessment in Histological Images (ATYPIA 2014) @ ICPR 2014, Stockholm, Sweden*
[3] *Collaborative Digital Pathology - FlexMIm (2013-2016) Funded by FUI - Consolidated Interministerial Fund: https://sites.google.com/site/telepathologiegroupeidf/home/composition-du-groupe-de-travail*
[4] *MIcroscope COgnitif : un explorateur visuel cognitif pour l'histopathologie. Application à la graduation du cancer du sein. – MICO : https://anr.fr/Projet-ANR-10-TECS-0015*

References:

1. *Ng SC et al. Lancet. 2018. doi: 10.1016/S0140-6736(17)32448-0.*
2. *Torres J et al. Lancet 2017. doi: 10.1016/S0140-6736(16)31711-1.*
3. *Pariente B et al. Gastroenterology 2015. doi: 10.1053/j.gastro.2014.09.015.*
4. *Furey TS et al. Nat Rev Gastroenter Hepatol. doi:10.1038/s41575-019-0118-x*
5. *Wu J et al. Am J Gastroenterol. 2019. doi: 10.14309/ajg.0000000000000237.*
6. *Chen P et al. Front Med 2020. doi: 10.3389/fmed.2020.00123.*
7. *Cleynen I et al. Lancet 2016. doi: 10.1016/S0140-6736(15)00465-1*
8. *Lee JC et al. Nat Genet. 2017. doi: 10.1038/ng.3755.*
9. *Alonso A et al. Gastroenterology. 2015. doi: 10.1053/j.gastro.2014.12.030.*
10. *Kugathasan S et al. Lancet. 2017. doi: 10.1016/S0140-6736(17)30317-3.*
11. *Hyams JS et al. Lancet. 2019. doi: 10.1016/S0140-6736(18)32592-3*
12. *Biasci D, et al. Gut 2019. doi:10.1136/gutjnl-2019-318343*
13. *Martin JC et al. Cell. 2019. doi: 10.1016/j.cell.2019.08.008.*
14. *Sokol H et al. Gut. 2019. doi: 10.1136/gutjnl-2019-318719.*
15. *Somineni HK. Gastroenterology. 2019. doi: 10.1053/j.gastro.2019.01.270.*
16. *Zemouri, R et al. Appl. Sci. 2019. doi:10.3390/app9081526.*
17. *Levine A et al. J Pediatr Gastroenterol Nutr. 2014. doi: 10.1097/MPG.0000000000000239.*
18. *Mehdi M et al. doi: 10.1136/annrheumdis-2019-eular.4703*
19. *Irshad, H et al. IEEE RBME, 7:97-114.2014, doi: 10.1109/RBME.2013.2295804*
20. *Zemouri, R. Appl. Sci.; doi:10.3390/app9081526.*
21. *Racoceanu D, Comput Med Imag Graphics 2015; doi: 10.1016/j.compmedimag.2014.09.004.*
22. *Ben Cheikh B et al. SPIE Medical Imaging 2017: https://doi.org/10.1117/12.2254452*
23. *Jiménez G and Racoceanu D. Front. Bioeng. Biotechnol. 2019, doi: 10.3389/fbioe.2019.00145.*
24. *Loménie N. and Racoceanu D, Pattern Recognition; doi: 10.1016/j.patcog.2012.01.021.*
25. *Racoceanu D. Pathobiology: J Immunopath Mol Cell Biol 2016. doi: 10.1159/000443964.*
26. *Chavannes M et al. J Crohns Colitis. 2019. doi: 10.1093/ecco-jcc/jjy206.*
27. *Cohen S et al. Pediatr Infect Dis J. 2017. doi: 10.1097/INF.0000000000001724.*
28. *Truffinet O et al. J Pediatr Gastroenterol Nutr. 2017. doi: 10.1097/MPG.0000000000001338.*
29. *Ouldani N et al. Inflamm Bowel Dis. 2018. doi: 10.1093/ibd/izy137.*
30. *Romagnoni A et al. Sci Rep. 2019. doi: 10.1038/s41598-019-46649-z.*
31. *Frémond ML et al. J Crohns Colitis. 2014. doi: 10.1016/j.crohns.2014.06.011.*
32. *Aguilar C et al. J Allergy Clin Immunol. 2014. doi: 10.1016/j.jaci.2014.04.031.*
33. *Thachil E et al. Gastroenterology. 2012. doi: 10.1053/j.gastro.2012.01.031.*
34. *Racoceanu, D., CMIG, 2: 2-15, 2015; doi: 10.1016/j.compmedimag.2014.09.004.*
35. *Vahabi, N et al. Unsupervised Multi-Omics Data Integration Methods: A Comprehensive Review, 2022; DOI=10.3389/fgene.2022.854752.*