Chaire de professeur junior Inria

[English version below, p. 6]

Établissement/organisme porteur : Centre Inria de l'Université de Lille
Nom du chef d'établissement/d'organisme : Mireille REGNIER
Site concerné : *ISITE-ULNE*
Région académique : Hauts de France

Établissements/organismes partenaires : *CHU Lille, Inserm, Université de Lille*

Nom du projet :
ENDOtyper les Maladies Inflammatoires Chroniques : comment obtenir une représentation unifiée des patients à partir de données hétérogènes ?
Acronyme : ENDOMIC

Thématique scientifique : santé numérique

Mots-clés :
Maladies autoimmunes ; endotypes ; hétérogénéité clinique ; analyse non supervisée ; science des données

Durée visée : 6 ans

Environnement financier : 1 070k € pour la durée du projet

Section (s) CNU/CoNRS/CSS correspondante (s) : CNU 26/27, CoNRS 6/41- CSS 5/6

Contact : cpj-lille@inria.fr

Stratégie d'établissement :

Le Numérique dans les Sciences du Vivant et la Santé est un domaine de recherche et d'innovation majeur d'Inria, auquel le récent programme de création d'équipes communes avec Inserm donne une nouvelle impulsion. Un défi stratégique est la structuration et le traitement des données de santé massives issues des hôpitaux en assurant l'interopérabilité et en préservant la confidentialité. Cet axe scientifique national d'Inria est cohérent avec la politique de l'Isite de Lille. Le CHU de Lille a mis en place un entrepôt de données de santé (EDS) institutionnel **INCLUDE** (INtegration Center of the Lille University hospital for Data Exploration) dont Inria Lille Nord Europe et l'Université de Lille sont partenaires. Cet EDS (dont le Professeur Vincent Sobanski est co-fondateur) rassemble 1,6 million de patients reçus au CHU de Lille depuis 2008 (autorisation CNIL post-RGPD obtenue en 2019). INCLUDE s'inscrit dans la dynamique nationale du *Health Data Hub* consistant à mettre en place des **Hubs régionaux** rassemblant les producteurs de données et les compétences d'exploitation. Un atout majeur est l'appartenance au réseau Européen EHDEN (European Health Data & Evidence Network – harmonisation des sources de données au format interopérable OMOP).

Stratégie du laboratoire d'accueil :

Le centre Inria et l'Université de Lille (UMRs Cristal et Paul Painlevé) développent des applications en santé avec le CHU de Lille (UMR INFINITE) dans une étroite collaboration qui s'est intensifiée en 2020. Ce projet vise à **renforcer et affirmer l'axe « Santé Numérique »** comme structurant et stratégique pour le centre Inria Lille – Nord Europe. Un objectif à court terme est la création possible d'une **nouvelle équipe-projet Inria-INSERM-CHU-Université de Lille**. Il s'intègre dans les hubs d'excellence « Santé de précision » et « Transition numérique au service de l'humain » de l'I-site Université Lille – Nord Europe. ENDOMIC relève du programme de recherche **« cluster PHENOMIX » de l'Isite** qui vise à développer un ensemble d'outils, de modèles et de composants logiciels permettant d'identifier des groupes de patients homogènes à partir des données massives. Il est **cohérent avec les projets scientifiques** d'équipes communes d'Inria et d'Université de Lille : parcours patient (UMR Paul Painlevé), protection des données de santé (Cristal) incluant l'Action Exploratoire FLAMED portée par Aurélien Bellet en collaboration avec INCLUDE sur le calcul fédéré au sein de réseaux hospitaliers (2 publications communes[1]) et interface homme-machine (Cristal).

[1] doi : 10.3233/SHTI200710 et 10.3233/SHTI210134

Résumé du projet scientifique :

Les maladies inflammatoires chroniques (MIC) sont caractérisées par une importante **hétérogénéité clinique** qui complexifie la prise en charge des patients. Concept clé de la médecine personnalisée, un “endotype” est un sous-type de maladie, défini par un mécanisme moléculaire ou une réponse particulière à un traitement. L’équipe “ENDOMIC” portée par le Pr Vincent Sobanski (Université et CHU de Lille, membre IUF Junior) vise à révéler des endotypes de MIC par une analyse multidimensionnelle intégrative de patients hautement phénotypés. Elle est basée sur leurs données cliniques et de suivi (informations de faible dimension) et sur des profils multi-omiques générés spécifiquement pour ce projet (informations de haute dimension). Retenu parmi les cinq lauréats nationaux de l’AMI “équipe-projet Inserm-Inria Santé numérique 2020”, ce projet constitue une avancée majeure vers une classification innovante car il intègre un très grand nombre de variables en prenant en compte la **dimension temporelle** et vise à concevoir des outils de **validation séquentielle** pour vérifier la pertinence d'une telle classification intégrative. Une originalité importante de ce projet est de travailler à rendre les endotypes intelligibles et exploitables en développant une **interface homme-machine** permettant aux acteurs (chercheurs, praticiens et même patients) de s'approprier cette nouvelle façon de classifier une maladie. La relation soignant-patient sera profondément modifiée, ce qui fera l’objet d’une évaluation approfondie.

Résumé du projet d’enseignement :

Le candidat participera aux enseignements liés à la **santé de précision** (Hubs santé et numérique de l’I-SITE ULNE dans le cadre du projet d’Établissement Public Expérimental) en y apportant son **expertise spécifique** (sciences des données, intelligence artificielle, analyse non supervisée). Ce programme gradué « santé de précision » est ouvert aux biologistes, médecins et pharmaciens, ainsi qu’aux (bio)informaticiens d’origine nationale et internationale, afin de **former la prochaine génération de leaders** de la recherche pour le monde universitaire, l’industrie et le monde des biotechnologies. Il comprend un parcours de Master en Biologie et Santé (M2) et un programme doctoral, **ouvert à toute la communauté scientifique**. Il est structuré autour de séminaires interactifs sur les technologies et les approches de pointe actuelles (par exemple les données en grand nombre, les OMICS, le séquençage de nouvelle génération, la biologie des systèmes et la bio-informatique intégrée, l’IA, la modélisation mathématique des systèmes biologiques, les thérapies innovantes, la gestion des cohortes, l’économie et l’éthique de la santé). Il pourra y développer un enseignement spécifique ou intégrer l’équipe pédagogique afin de structurer et animer cette **offre de formation stratégique pour l’EPE**.

Diffusion scientifique :
Le titulaire de la chaire sera tenu de valoriser les résultats obtenus au sein de publications dans des **journaux scientifiques** reconnus (cible rang A) et dans les **congrès internationaux** de référence pour les disciplines informatique et santé. Le caractère transdisciplinaire de la recherche menée amènera à la publication d'articles dans des **revues généralistes à forte audience** (core medical journals) et à la rédaction de chapitres ou ouvrages pédagogiques. Il devra aussi avoir une appétence certaine pour la **vulgarisation scientifique** en particulier auprès du grand public (journées portes ouvertes, lycéens, fête de la science).

Science ouverte :
L'interface homme-machine permettra la **navigation dans la représentation virtuelle des endotypes**. Les professionnels de santé pourront saisir certaines variables dans un module de requête qui leur permettra de prédire l'affectation potentielle d'un patient donné dans les endotypes. En outre, ils pourront implémenter leurs données sous la forme d'une **base de données interopérable** qui pourra être intégrée à l'algorithme et ainsi enrichir la définition des endotypes (**open science**). Les patients et les professionnels de santé seront interrogés sur leurs **besoins**. Ils seront impliqués dans des **activités de co-création** pour finalement obtenir des spécifications répondant à leurs besoins et attentes. Les spécifications de cette interface intégreront également la possibilité pour un professionnel de santé de positionner un patient donné dans les endotypes à partir de quelques variables entrées dans le modèle, l'interface fournissant un **degré de confiance pour cette classification** (projection probabiliste). L'algorithme devra également être capable de suggérer au professionnel de santé les examens à réaliser (variables supplémentaires) pour améliorer la classification de son patient, voire orienter sa prise en charge, comme le propose l'algorithme DETECT mis en oeuvre il y a quelques années pour la décision de réaliser un cathétérisme cardiaque droit chez les patients atteints de sclérodermie systémique suspectés d'hypertension pulmonaire.
Le travail réalisé sur l'interopérabilité des données au format OMOP au sein du CHU de Lille permettra à l'outil d'être ouvert « by design ». Les données nécessaires au développement de l'algorithme seront disponibles selon les **principes de la science ouverte**, en accord avec nos publications récentes[2] (nous sommes membres de l'initiative EHDEN (European Health Data & Evidence Network).

Science et société :
L'appropriation de ce nouvel outil par les acteurs sera assurée par des publications dans des **revues en accès libre**, la participation à des **congrès scientifiques et de patients**, la **documentation sur le site internet**, la communication sur les **réseaux sociaux** avec les professionnels et les associations de patients (par exemple dans les réseaux dans lesquels nous sommes déjà impliqués : FAI2R, EUSTAR) et des **séminaires scientifiques**.
Nous sommes très investis dans la **communication scientifique, en particulier auprès du grand public** (le Pr Vincent Sobanski est Vice-Doyen Communication de l'UFR des Sciences de Santé et du Sport de l'Université de Lille). La Taskforce de l'I-SITE ULNE a initié un cycle de conférences grand public sur la Covid-19. Une plateforme YouTube a été mise en place[3], dont les vidéos ont été vues 27 000 fois et largement relayée dans les médias et labélisée par le Ministère de l'Enseignement Supérieur et de la Recherche.

[2] doi : 10.3233/SHTI200710 et 10.3233/SHTI210134
[3] https://www.youtube.com/channel/UCKv7KQLa0w2AVq6hzcY3q1w/featured

Indicateurs :

Indicateurs prioritaires	
Précoces (<2 ans)	1. Recrutement des ingénieurs, doctorants et post-doctorants constituant la chaire 2. Participation à la soumission de projets « santé » d'envergure nationale ou internationale (ANR, ERC, réseau européen…) dans la thématique de l'équipe ENDOMIC
Moyen terme (2-4 ans)	1. Publications dans des journaux scientifiques de haut niveau 2. Mise en place d'un enseignement spécifique type DU ou certificat programme gradué (structuration par le titulaire de la chaire) 3. Dépôt par le titulaire de la chaire de projets « informatique » d'envergure nationale ou internationale (ANR, ERC, réseau européen…) dans la thématique de l'équipe ENDOMIC
Long terme (4-6 ans)	1. Dynamique de science ouverte (actions dans le cadre d'EHDEN et de l'AEx FLAMED, mise à disposition de jeux de données et codes publics, intégration de nouvelles données FAIR dans l'algorithme de classification intégrative) 2. Transposition de l'outil de classification vers d'autres champs médicaux (nouvelles collaborations, constitution d'une start-up)
Indicateurs secondaires	
Précoces (<2 ans)	1. Publications dans des revues informatiques ou médicales de niveau intermédiaire 2. Participation aux congrès internationaux des sociétés savantes
Moyen terme (2-4 ans)	1. Intégration dans des réseaux de recherche notamment européens (visibilité internationale) 2. Organisation d'un séminaire scientifique international
Long terme (4-6 ans)	1. Collaborations académiques en particulier capacité de l'équipe ENDOMIC à envoyer/recevoir des étudiants/chercheurs en mobilité internationale 2. Mise en place de contrats partenariaux

Inria Chair of Junior Professor

Supporting institution/organization: Inria center at the University of Lille
Head of the institution/organization: Mireille REGNIER
Site concerned: *ISITE-ULNE*
Academic Region: Hauts de France

Partner institutions/organizations: *CHU Lille, Inserm, Université de Lille*

Project name:
ENDOtyping Chronic Inflammatory Diseases: how to obtain a unified representation of patients from heterogeneous data?
Acronym: ENDOMIC

Scientific topic: digital health

Keywords:
Autoimmune diseases; endotypes: clinical heterogeneity; unsupervised analysis; data science

Target duration: 6 years

Financial overview: 1 070k € for the project

Section (s) CNU/CoNRS/CSS corresponding: CNU 26/27, CoNRS 6/41- CSS 5/6

Contact : cpj-lille@inria.fr

Establishment strategy:

Digital Life Sciences and Health is a major area of research and innovation at Inria, and the recent program to create joint teams with Inserm is giving it new impetus. A strategic challenge is the structuring and processing of massive health data from hospitals while ensuring interoperability and preserving confidentiality. This national scientific axis of Inria is consistent with the policy of the Lille Isite. The Lille University Hospital has set up an institutional clinical data warehouse (CDW) called **INCLUDE** (INtegration Center of the Lille University hospital for Data Exploration), in which Inria Lille Nord Europe and the University of Lille are partners. This CDW (of which Professor Vincent Sobanski is a co-founder) gathers 1.6 million patients received at the Lille University Hospital since 2008 (CNIL GDPR-compliant authorization obtained in 2019). INCLUDE is part of the national dynamic of the *Health Data* Hub consisting of setting up **regional Hubs** bringing together data producers and operating skills. A major asset is its membership in the European Health Data & Evidence Network (EHDEN), which harmonizes data sources in OMOP interoperable format.

Strategy of the host laboratory:

The Inria center and the University of Lille (UMRs Cristal and Paul Painlevé) are developing health applications with the University Hospital of Lille (UMR INFINITE) in a close collaboration that has intensified in 2020. This project aims to **strengthen and affirm the "Digital Health" axis** as a structuring and strategic axis for the Inria Lille - Nord Europe center. A short-term objective is the **creation of a new Inria-INSERM-CHU-University of Lille project team**. It is integrated into the "Precision Health" and "Digital Transition for Humans" hubs of excellence of the Lille-Nord Europe University I-site. ENDOMIC is part of **the I-site "PHENOMIX cluster" research program**, which aims to develop a set of tools, models and software components to identify homogeneous groups of patients from massive data. It is **consistent with the scientific projects** of joint Inria and University of Lille teams: patient pathways (UMR Paul Painlevé), health data protection (Cristal) including the FLAMED Exploratory Action led by Aurélien Bellet in collaboration with INCLUDE on federated computing within hospital networks (2 joint publications) and man-machine interface (Cristal).

Summary of the scientific project:

Chronic inflammatory diseases (CID) are characterized by significant **clinical heterogeneity** which complicates patient management. A key concept in personalized medicine is "endotype" as a disease subtype defined by a molecular mechanism or a particular response to treatment. The "ENDOMIC" team led by Prof. Vincent Sobanski (Lille University and University Hospital of Lille, Junior IUF member) aims at revealing endotypes of CID through an integrative multidimensional analysis of highly phenotyped patients. It is based on their clinical and follow-up data (low dimensional information) and on multi-omics profiles generated specifically for this project (high dimensional information). Selected among the five national winners of the AMI "Inserm-Inria Digital Health 2020 project-team", this project constitutes a major advance towards an innovative classification because it integrates a very large number of variables by taking into account **the temporal dimension** and aims to design **sequential validation** tools to verify the relevance of such an integrative classification. An important originality of this project is to work on making the endotypes intelligible and exploitable by developing a **man-machine interface** allowing the actors (researchers, practitioners and even patients) to appropriate this new way of classifying a disease. The relationship between caregiver and patient will be profoundly modified, which will be the subject of an in-depth evaluation.

Summary of the teaching project:

The candidate will participate in the teaching related to **precision health** (Health and Digital Hubs of the ULNE I-SITE in the framework of the Experimental Public Establishment project) by bringing his/her **specific expertise** (data science, artificial intelligence, unsupervised analysis). This "precision health" graduate program is open to biologists, physicians and pharmacists, as well as to (bio)computer scientists of national and international origin, in order to **train the next generation of research leaders** for academia, industry and the biotechnology world. It includes a Master's program in Biology and Health (M2) and a PhD program, **open to the entire scientific community**. It is structured around interactive seminars on current cutting-edge technologies and approaches (e.g., big data, OMICS, next-generation sequencing, systems biology and integrated bioinformatics, AI, mathematical modeling of biological systems, innovative therapies, cohort management, health economics and ethics). He/she will be able to develop specific teaching or join the teaching team in order to structure and lead this **strategic training offer for the University of Lille**.

Scientific diffusion:
The chairholder will be required to promote the results obtained through publications in **recognized scientific journals** (target rank A) and in **international conferences** of reference for the computer science and health disciplines. The transdisciplinary nature of the research conducted will lead to the publication of articles in **generalist journals with a large audience** (core medical journals) and to the writing of chapters or educational books. He/she should also have a certain appetite for **scientific popularization**, particularly with the general public (university open days, high school students, French science days).

Open Science:
The man-machine interface will allow **navigation in the virtual representation of endotypes**. Healthcare professionals will be able to enter certain variables in a query module that will allow them to predict the potential assignment of a given patient to the endotypes. In addition, they will be able to implement their data in the form of an **interoperable database** that can be integrated into the algorithm and thus enrich the definition of endotypes (**open science**). Patients and health professionals will be asked about their **needs**. They will be involved in **co-creation activities** to finally obtain specifications that meet their needs and expectations. The specifications of this interface will also integrate the possibility for a health professional to position a given patient in the endotypes from a few variables entered in the model, the interface providing a **degree of confidence for this classification** (probabilistic projection). The algorithm should also be able to suggest to the healthcare professional the examinations to be performed (additional variables) to improve the classification of his patient, or even to guide his management, as proposed by the DETECT algorithm implemented a few years ago for the decision to perform a right heart catheterization in patients with systemic scleroderma suspected of having pulmonary hypertension.
The work done on the interoperability of data in OMOP format within the Lille University Hospital will allow the tool to be open "by design". The data necessary for the development of the algorithm will be available according to the **principles of open science**, in accordance with our recent publications (we are members of the EHDEN initiative (European Health Data & Evidence Network).

Science and society:
The appropriation of this new tool by the actors will be ensured by publications in **open access journals**, participation **in scientific and patient congresses**, **documentation on the website**, communication on **social networks** with professionals and patient associations (e.g. in the networks in which we are already involved: FAI2R, EUSTAR) and **scientific seminars**.
We are very involved in **scientific communication, especially with the general public** (Prof. Vincent Sobanski is Vice-Dean of Communication at the University of Lille's Health and Sports Sciences Department). The ULNE I-SITE Taskforce has initiated a series of conferences for the general public on Covid-19. A YouTube platform has been set up, whose videos have been seen 27,000 times and widely relayed in the media and labelled by the Ministry of Higher Education and Research.

Indicators:

Primary indicators	
Short term (<2 years)	1. Recruitment of engineers, doctoral students and post-doctoral fellows constituting the Chair 2. Participation in the submission of national or international "health" projects (ANR, ERC, European network...) in the field of ENDOMIC
Mid term (2-4 years)	1. Publications in high level scientific journals 2. Establishment of a specific teaching program such as a university diploma or certificate graduate program (structured by the chairholder) 3. Submission by the chairholder of "computer science" projects of national or international scope (ANR, ERC, European network...) in the thematic of the ENDOMIC team
Long term (4-6 years)	1. Open science dynamic (actions in the framework of EHDEN and AEx FLAMED, making public datasets and codes available, integration of new FAIR data in the integrative classification algorithm) 2. Transposition of the classification tool to other medical fields (new collaborations, creation of a start-up)
Secondary indicators	
Short term (<2 years)	1. Publications in intermediate level computer or medical journals 2. Participation in international conferences of scientific societies
Mid term (2-4 years)	1. Integration into research networks, particularly European ones (international visibility) 2. Organization of an international scientific seminar
Long term (4-6 years)	1. Academic collaborations, in particular capacity of the ENDOMIC team to send/receive students/researchers in international mobility 2. Establishment of partnership contracts