Ecole doctorale Environnements-Santé
Dossier de projet de thèse « Contrat doctoral Etablissements »
ANNEE 2026

Machine Learning Framework for Multimodal Analysis of Neurodevelopmental and Neurodegenerative Mechanisms in iPSC-Derived Brain Organoids

1) Renseignements administratifs sur la direction de thèse[1] (1 page maximum) :

Directeur de thèse HDR :
Nom : Thauvin
Prénom : Christel
Section CNU : 4704 - Génétique
Grade : PUPH Classe exceptionnelle
HDR : Date de soutenance……10 octobre 2012…….. Discipline : ……Génétique……
l'HDR devra être soutenue, ou sa soutenance autorisée, au moment du dépôt du présent projet.
Coordonnées (adresse, courriel, téléphone) : équipe GAD, 15 bd Maréchal de Lattre de Tassigny, 21000 Dijon ; christel.thauvin@inserm.fr ; 03.80.39.66.59
Unité d'appartenance (intitulé, label, n°, directeur) : Inserm-UBE 1231 CTM Pr F. Ghiringhelli

Co-directeur de thèse éventuel :
Nom : Jego
Prénom : Laurence
Section CNU : *65 – Biologie cellulaire*
Grade : Ingénieur de recherche
HDR : non ☐ ; oui X Date de soutenance : 19/12/2017. Discipline : Environnement / santé.
Coordonnées (adresse, courriel, téléphone) : équipe GAD, 15 bd Maréchal de Lattre de Tassigny, 21000 Dijon ; laurence.duplomb@chu-dijon.fr ; 03.80.39.66.59
Unité d'appartenance (intitulé, label, n°, directeur) : Inserm-UBE 1231 CTM Pr F. Ghiringhelli

2) Descriptif du projet de thèse (devra inclure les rubriques suivantes) :

- nom et label de l'unité de recherche (ainsi que l'équipe interne s'il y a lieu)
- localisation
- nom du directeur de thèse et du co-directeur s'il y a lieu
- adresse courriel du contact scientifique
- titre du projet
- description du projet (2 pages maximum)
- Financement du projet – partie Recherche (montants acquis, type de contrat)
- connaissances et compétences requises

[1] ATTENTION : selon l'article 16 de l'arrêté du 25 mai 2016, le total d'encadrants ne peut pas dépasser 2, sauf si l'un des encadrants appartient au monde socio-économique, qui peut venir en sus, ou en cas de co-tutelle; Le décompte des co-encadrements se fera au prorata du nombre d'encadrants : 1 pour 1 encadrant, ½ pour deux encadrants.

Nom et label de l'unité de recherche (ainsi que l'équipe interne s'il y a lieu) : CTM Inserm U1231, GAD Team, Université de Bourgogne Europe

Localisation : Université de Bourgogne, Batiment B3

Directeur de thèse : Pr Christel Thauvin

co-directeur de thèse : Dr Laurence Duplomb-Jego

Contact scientifique :
christel.thauvin@chu-dijon.fr OR laurence.duplomb@chu-dijon.fr

Financement du projet : Thèse : Bourse MRT. Matériels et échantillons : Projet INSERM CHAIR (ANR-23-CPJ1-0133-01, financement déjà acquis, 200k€)

Connaissances et compétences requises : Le candidat idéal est titulaire d'un Master en bioinformatique, biologie computationnelle ou dans un domaine connexe, avec une expérience démontrée en analyse de données transcriptomiques (par exemple en R, Python, etc.). Une familiarité avec les bases de données orientées graphe, l'apprentissage automatique ou les pipelines d'analyse d'images constitue un atout. Une motivation à travailler à l'interface entre neurosciences computationnelles et expérimentales, ainsi qu'à collaborer étroitement avec des équipes de laboratoire pour co-développer des pipelines biologiquement pertinents, est indispensable.

Résumé en français et anglais (limité chacun à 1800 caractères)

FR: Les troubles neurodéveloppementaux sont des affections hétérogènes qui touchent des millions de personnes et représentent un lourd fardeau sociétal. De manière frappante, certains patients développent un parkinsonisme ou une démence précoces, soulignant un continuum entre troubles du développement et maladies neurodégénératives. Des données émergentes suggèrent que ces dernières peuvent trouver leur origine ou être aggravées par des anomalies développementales, les variants génétiques rares jouant un rôle majeur. Les organoïdes cérébraux dérivés d'iPSC offrent une plateforme innovante pour modéliser ces pathologies, en reproduisant l'architecture, la diversité cellulaire et les dynamiques transcriptionnelles du cerveau. Notre équipe a montré que la mutation rare Miro1 p.R272Q est causale dans la maladie de Parkinson, et nous étendons désormais cette expertise à d'autres variants pathologiques tels que PTBP1, SOGA1 et ANKRD11.
Ce projet postule que les mutations rares génèrent des signatures transcriptomiques et morphométriques convergentes dans les organoïdes, recoupant des mécanismes de neurodégénérescence. Pour tester cette hypothèse, nous établirons un pipeline scRNA-seq robuste, développerons des workflows automatisés d'imagerie, et intégrerons ces données multimodales grâce à des approches avancées d'apprentissage automatique. L'intégration des données transcriptomiques et morphologiques avec des ressources publiques permettra d'identifier des voies moléculaires communes, de définir des sous-groupes homogènes de patients et de proposer des hypothèses fonctionnelles. Ce travail fournira un cadre intégré pour l'interprétation des pathologies liées aux mutations rares, des résultats biologiquement interprétables pour guider les équipes expérimentales, et contribuera à des stratégies de médecine de précision applicables à divers systèmes organoïdes.

EN: Neurodevelopmental disorders are highly heterogeneous conditions that affect millions worldwide and impose a major societal burden. Strikingly, some patients develop early parkinsonism and dementia, underscoring a continuum between developmental and degenerative disorders. Emerging evidence suggests that neurodegenerative diseases may

originate from, or be exacerbated by, developmental abnormalities, with rare genetic variants playing a critical role. Human iPSC-derived brain organoids provide a transformative platform to model these disorders, recapitulating brain architecture, cellular diversity, and transcriptional dynamics. Our team has shown that the rare Miro1 p.R272Q mutation is causative of Parkinson's disease, and we now extend this expertise to other pathological variants such as PTBP1, SOGA1, and ANKRD11.
This project hypothesizes that rare mutations generate convergent transcriptomic and morphometric signatures in organoids, overlapping with mechanisms of neurodegeneration. To test this, we will establish a robust scRNA-seq pipeline, develop automated imaging workflows, and integrate multimodal datasets using advanced machine learning approaches. By combining transcriptomic and imaging data with public resources, we aim to uncover shared molecular pathways, identify biologically homogeneous patient subgroups, and propose data-driven hypotheses for functional validation. Ultimately, this work will deliver an integrated framework for interpreting mutation-driven pathologies, generate biologically interpretable results to guide experimental teams, and contribute to precision medicine strategies applicable across organoid systems.

Préciser le domaine de compétence dans la liste ci-dessous :
2 choix possibles maximum –les intitulés sont ceux d'ABG),
Ex : vous ne pouvez pas sélectionner uniquement informatique lorsqu'il est indiqué informatique, électronique **:**

- Informatique, électronique
- Santé, médecine humaine, vétérinaire

Title : Machine Learning Framework for Multimodal Analysis of Neurodevelopmental and Neurodegenerative Mechanisms in iPSC-Derived Brain Organoids

Indiquez quelques mots clés concernant votre sujet) : Neurodevelopmental disorders; neurodegeneration; iPSC-derived brain organoids; scRNA-seq; high-content imaging; machine learning; rare genetic variants; multimodal integration; precision medicine

Background: Neurodevelopmental disorders are highly heterogeneous conditions that disrupt brain growth and function, leading to lifelong difficulties in cognition, communication, and behavior (Gao et al., 2025, Med Rev). Their burden is immense: over 1,700 genes have been implicated, affecting nearly 4 million individuals in France alone, with an annual societal cost of ~28 billion euros (Gidziela et al., 2023, Nat Hum Behav & Schoentgen et al., 2025, Encephale). Patients often present with multiple comorbidities, and strikingly, some develop early parkinsonism and dementia (von Scheibler et al., 2023, Mov Disord Clin Pract), highlighting a continuum between developmental and degenerative disorders. In fact, emerging evidence indicates that neurodegenerative diseases may originate from, or be worsened by, developmental abnormalities, as shown by clinical, genomic, and molecular similarities (Hickman et al., 2022, Nat Rev Neurol). In addition, rare genetic variants have been found to play a major role in neurodegenerative conditions (Makarious et al., 2023, Brain).

Human iPSC-derived brain organoids now provide a transformative platform to model these disorders *in vitro*. They recapitulate architecture, cellular diversity, and transcriptional dynamics of the brain, enabling pathological validation and mechanistic studies of rare mutations (Eichmüller & Knoblich, 2022, Nat Rev Neurol; Cervo et al., 2021, Nat Rev Neurol). Our team has already demonstrated that the rare Miro1 p.R272Q is causative of PD, since it leads to the loss of dopaminergic neurons in aged knock-in mice and patient-specific midbrain organoids due to mitochondrial dysfunction and altered developmental

differentiaton (Chemla et al., 2025, Brain & Zagare et al., 2025, NPJ Syst Biol Appl). Building on this expertise, we are expanding to other pathological rare mutations such as PTBP1, SOGA1, and ANKRD11.

Single-cell RNA sequencing (scRNA-seq) has revolutionized organoid research, enabling precise characterization of cell populations, developmental trajectories, and mutation-driven molecular changes (Tanay & Regev, 2017, *Nature*). High-content imaging complements this by providing quantitative descriptors of tissue organization, morphology, proliferation, stratification, and neuronal maturation. When combined with other omics layers (genomics, metabolomics, proteomics), these approaches offer unprecedented resolution of disease mechanisms. Despite their power, these datasets are still analyzed largely in isolation, limiting our ability to fully capture the complexity of mutation-driven pathologies.

Approaches for integrating heterogeneous data have expanded considerably in recent years, particularly in the context of single-cell multi-omics datasets (Hao et al., 2021, *Cell*). Methods based on graphs, deep learning, or variational models (MOFA+, Argelaguet et al., 2020, *Mol. Syst. Biol.*) enable the identification of common factors across different axes of analysis. In parallel, automated analysis of biological images has benefited from advances in deep learning, especially convolutional neural networks (CNNs) for segmentation and extraction of morphological descriptors (Ronneberger et al., 2015, U-Net). The extraction of quantitative features (size, shape, intensity, spatial organization) allows images to be transformed into variables that can be exploited by statistical or predictive models. By extracting integrating imaging and transcriptomic data with public datasets (e.g., Reactome, DisGeNet, DGIdb, IntAct), we will uncover novel convergent pathways and biologically homogeneous subgroups within patients with rare mutations.

Altogether, the central innovation of this project is the **systematic integration of single-cell transcriptomic and quantitative imaging data** from brain organoids carrying rare pathogenic mutations. Using state-of-the-art machine learning, we aim to identify convergent transcriptomic and morphometric signatures, map them onto disease-relevant pathways, and relate them to patient heterogeneity.

Hypothesis: We hypothesize that rare genetic mutations underlying neurodevelopmental disorders generate convergent transcriptomic and morphometric signatures in iPSC-derived brain organoids, which predispose to or overlap with mechanisms of neurodegeneration. By systematically integrating scRNA-seq and quantitative imaging data through advanced machine learning, we will uncover shared molecular pathways and define biologically homogeneous patient subgroups, providing new mechanistic insights and guiding precision medicine strategies.

To pursue our hypothesis, we intend to address the following **main objectives**:

1. **Establish a robust scRNA-seq analysis pipeline tailored to brain organoids** (quality control, cell annotation, trajectory inference, differential analyses).
2. **Develop an automated analysis workflow for immunofluorescence images** (segmentation, feature extraction, normalization).
3. **Build an integrative analysis model based on machine learning approaches** (shared components, graph-based methods, supervised/unsupervised learning) applicable to mutation-driven pathologies.
4. **Identify convergent molecular pathways across rare mutations that bridge neurodevelopmental and neurodegenerative mechanisms, and propose data-driven hypotheses for functional validation.**

Workplan: In the first year, the doctoral candidate will establish the analytical foundations by conducting a literature review on brain organoids and neurodevelopmental disorders, and by developing a reproducible scRNA-seq pipeline using existing datasets. This will

include cell annotation, comparison with reference atlases, and initial differential analyses, alongside training in imaging techniques and defining requirements for image analysis. The main deliverables will be a validated transcriptomic pipeline and a first comparative pilot study.

The second year will focus on imaging quantification and the initial integration of multimodal data. A segmentation pipeline will be developed to extract morphometric and spatial descriptors, which will then be harmonized with transcriptomic datasets. An initial integrative model will be implemented, and preliminary results will be presented at conferences. Deliverables will include an integrated database and the submission of a methodological or application-focused article.

In the third year, advanced modeling approaches will be refined to identify convergent signatures across mutations, explore patient stratification, and validate findings in collaboration with the experimental team. This final phase will culminate in publications and the doctoral thesis, with deliverables including at least one major publication and a reproducible model generalizable to other pathological contexts.

Expected Results: This project will deliver an integrated framework to interpret molecular and morphological alterations driven by rare pathological mutations. The work will further uncover convergent pathways that bridge neurodevelopmental and neurodegenerative mechanisms, providing new mechanistic insights, which can lead to clinical translation through discover of biomarkers, development of new treatment and patient stratification. In the longer term, the analytical approach will be transferable to other experimental systems contributing to precision medicine through the integration of complex multimodal omics data.

References:

Gao S, Shan C, Zhang R, Wang T. (2025). Genetic advances in neurodevelopmental disorders. Medical review, 5(2):139–51. https://doi.org/10.1515/mr-2024-0040

Gidziela A, Ahmadzadeh YI, Michelini G, Allegrini AG, Agnew-Blais J, Lau LY, et al. (2023) A meta-analysis of genetic effects associated with neurodevelopmental disorders and co-occurring conditions. Nat Hum Behav, 7(4):642–56. https://doi.org/10.1038/s41562-023-01530-y

Schoentgen B, Chatin E, Sallé H, Leroy M, Bonnet N, Défontaines B. (2025).[Socioeconomic impact of neurodevelopmental disorders in France]. Encephale, 51(5):557–62. https://doi.org/10.1016/j.encep.2024.11.015

Hickman, R. A., O'Shea, S. A., Mehler, M. F., & Chung, W. K. (2022). Neurogenetic disorders across the lifespan: from aberrant development to degeneration. *Nature Reviews. Neurology*, *18*(2), 117–124. https://doi.org/10.1038/s41582-021-00595-5

Makarious, M. B., Lake, J., Pitz, V., Fu, A. Y., Guidubaldi, J. L., Solsberg, C. W., et el. (2023). Large-scale rare variant burden testing in Parkinson's disease. *Brain*, awad214. https://doi.org/10.1093/brain/awad214

Eichmüller, O. L., & Knoblich, J. A. (2022). Human cerebral organoids - a new tool for clinical neurology research. *Nature Reviews. Neurology*, *18*(11), 661–680. https://doi.org/10.1038/s41582-022-00723-9

Rivetti di Val Cervo, P., Besusso, D., Conforti, P., & Cattaneo, E. (2021). hiPSCs for predictive modelling of neurodegenerative diseases: dreaming the possible. *Nature Reviews. Neurology*, *17*(6), 381–392. https://doi.org/10.1038/s41582-021-00465-0

Chemla A, Arena G, Sacripanti G, Barmpa K, Zagare A, Garcia P, et al. (2025). Parkinson's disease mutant Miro1 causes mitochondrial dysfunction and dopaminergic neuron loss. Brain, 148:10:3607–3622. https://dx.doi.org/10.1093/brain/awaf051

Zagare A, Sauter T, Barmpa K, Pacheco M, Krüger R, Schwamborn JC, Saraiva C (2025). MIRO1 mutation leads to metabolic maladaptation resulting in Parkinson's disease-associated dopaminergic neuron loss. npj Systems Biology and Applications, 11:1. https://doi.org/10.1038/s41540-025-00509-x

Tanay A., Regev A. (2017). Scaling single-cell genomics from phenomenology to mechanism. *Nature*, 541: 331–338. https://doi.org/10.1038/nature21350

Hao Y., Hao S., Andersen-Nissen E., et al. (2021). Integrated analysis of multimodal single-cell data. *Cell*, 184(13): 3573–3587.e29. https://doi.org/10.1016/j.cell.2021.04.048

Argelaguet R., Arnol D., Bredikhin D., et al. (2020). MOFA+: a statistical framework for comprehensive integration of multi-modal single-cell data. *Molecular Systems Biology*, 16(6): e9443. https://doi.org/10.15252/msb.20209443

Ronneberger O., Fischer P., Brox T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. In: *Medical Image Computing and Computer-Assisted Intervention (MICCAI 2015)*, LNCS 9351, Springer. https://doi.org/10.1007/978-3-319-24574-4_28