



## Intitulé de l'offre de stage

Using Machine Learning and Data Analysis methods on transcriptome data to predict from diagnosis the minimal residual disease (MRD) after treatment

<b>Stagiaire en</b>	Informatique (Deep Learning)
<b>Affectation</b>	Équipe commune CRCT/IRIT
<b>Durée</b>	5 à 6 mois
<b>Rémunération</b>	environ 1200 € net mensuel
<b>Date de la publication</b>	15/11/2022
<b>Date d'embauche prévue</b>	mars / avril 2023 suivant disponibilités
<b>Lieu</b>	CRCT, 2 Avenue Hubert Curien – 31100 Toulouse

## Le Centre de Recherche en Cancérologie de Toulouse

Le CRCT est une unité de recherche conjointe entre l'Inserm et l'Université Toulouse III Paul Sabatier. Au cœur de l'Oncopole de Toulouse, le CRCT, avec tous ses partenaires (institutionnels, universitaires, cliniques, industriels, caritatifs...), stimule l'innovation en termes de recherche et d'enseignement dans la lutte contre le cancer. Le CRCT conduit une approche intégrée entre la recherche, les soins et l'enseignement, dans une logique transversale et multidisciplinaire.

## L'Institut de Recherche en Informatique de Toulouse

L'Institut de Recherche en Informatique de Toulouse (IRIT), une des plus imposantes Unité Mixte de Recherche au niveau national, est l'un des piliers de la recherche en Occitanie avec ses 700 membres, permanents et non-permanents. De par son caractère multi-tutelle (CNRS, Universités toulousaines), son impact scientifique et ses interactions avec les autres domaines, le laboratoire constitue une des forces structurantes du paysage de l'informatique et de ses applications dans le monde du numérique, tant au niveau régional que national.

## L'équipe

L'informatique et l'intelligence artificielle sont appelés à prendre une place croissante dans le monde de la recherche médicale, et en particulier dans le monde de la recherche contre le cancer. L'équipe commune IRIT/CRCT, co-localisée sur le site de l'Oncopole de Toulouse, a pour but de faire travailler ensemble et sur le même site chercheurs en informatique, chercheurs dans le domaine du cancer et médecins. Elle peut s'appuyer sur le plateau technique du CRCT (analyses biologiques, séquençages génomique, . . .) et sur les moyens de calcul de la région Occitanie ([CALMIP](#)).

## Objet du stage

Multiple myeloma (MM) is a hematological malignancy that affects bone marrow plasma cells. It remains an incurable malignancy, despite impressive progresses in the past 10 years. Physicians are facing to a very heterogeneous disease, with survival ranging from few months to more than 20 years. This heterogeneity is supposed to be mainly related to the huge genetic variability observed in the malignant plasma cells. Minimal residual disease (MRD), which corresponds to the number of MM residual cells after treatment, has clearly demonstrated its powerful prognostic impact in MM (Munshi et al. Blood Adv 2020). In some studies, an undetectable MRD ( $<10^{-6}$ , meaning less than one MM cells among 1 000 000 analyzed bone marrow cells) is able of abrogating the initial prognosis for some high-risk patients (Perrot et al, Blood 2018, Goicoechea et al. Blood 2020). However, this information is becoming available too late in the patient management. From a risk-adjusted treatment perspective, we believe it is crucial to be able to accurately predict which patients require aggressive treatment and which do not (which also has a medico-economic interest in this era of sometimes extremely costly strategies). While the reasons why a patient does not achieve undetectable MRD are probably multiple, the profile of patients able of achieving it should be easier to identify.

The first step of this study will be to use data analysis, machine learning and deep learning methods to process the results of RNA-seq on 20000 genes for 600 patients measured before treatment with quadruplet based induction (Daratumumab Bortezomib Thalidomide Dexamethasone), and correlate these data with the rate of MRD at the end of treatment. Being able to properly predict MRD after treatment from prior RNA-seq analysis would be of an invaluable interest both regarding the patient quality of life, by using lighter treatments when possible, and regarding life expectancy, by using more intensive treatments for patient that present high probability of resistance.

## Formation

École d'ingénieur, de préférence avec spécialisation en informatique ou mathématiques (en année de césure ou stage long). Master 2 informatique ou mathématiques appliquées.

## Compétences attendues

Des connaissances en science des données, apprentissage et réseaux de neurones seront appréciées.

## Modalité de candidature

- > CV à envoyer à: [myelome@stages-medecine-numerique.fr](mailto:myelome@stages-medecine-numerique.fr)
- > Date limite de candidature : 31/01/2023
- > Encadrants: Professeur Hervé Avet-Loiseau (CHU de Toulouse/IUCT-Oncopole/CRCT), Professeur Jill Corre (CHU de Toulouse/IUCT-Oncopole/CRCT), David Simoncini (IRIT/CRCT), Sandrine Mouysset (IRIT/CRCT), Josph Gergaud (IRIT/CRCT), Sylvain Cussat-Blanc (IRIT/CRCT)

Une première sélection sera effectuée sur la base des CVs reçus. Les candidats seront immédiatement informés du résultat, et ceux qui seront retenus à l'issue de la première sélection seront invités à un ou deux entretiens individuels en téléconférence avant sélection définitive.

L'ensemble des offres de stage est disponible sur <http://www.stages-medecine-numerique.fr>.