



## Intitulé de l'offre de stage

Identification par machine learning des facteurs de risques cliniques, génomiques (DNaseq) et phénotypiques (RNASeq) de 1.357 patients de 18 à 60 ans atteints de leucémie myéloïde aiguë traités par chimiothérapie intensive dans un unique protocole clinique.

<b>Stagiaire en</b>	Informatique (Machine Learning)
<b>Affectation</b>	Équipe commune CRCT/IRIT
<b>Durée</b>	5 à 6 mois
<b>Rémunération</b>	environ 670 € net mensuel
<b>Date de la publication</b>	1/12/2023
<b>Date d'embauche prévue</b>	mars / avril 2024 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

La leucémie aiguë myéloïde (LAM) est une tumeur maligne hématopoïétique fréquente chez les adultes, associée à un mauvais pronostic. La pathogenèse moléculaire de la LAM est caractérisée par l'acquisition d'un petit nombre de mutations conduisant à une architecture clonale complexe. Les résultats génétiques limités combinés au caryotype sont utilisés pour définir le pronostic de la maladie et construire le projet clinique du patient. Le génotype moléculaire seul a démontré son utilité pour la prise en charge individuelle des patients atteints de LAM mais, actuellement, seuls quelques gènes de cancer ont un traitement adapté, la fréquence des mutations pour chaque gène de cancer est très variable et chaque LAM a typiquement plusieurs mutations drivers conduisant à un nombre très élevé de combinaisons de mutations.

Nous voulons augmenter la prédiction de l'évolution des jeunes LAM traitées avec le nouveau standard de traitement en France et développer un support solide pour choisir parmi différentes options thérapeutiques (thérapies ciblées, allogreffe ou thérapie de maintenance) combinant le génotype de la maladie à son phénotype.

Le stage a pour finalité d'analyser les données acquises concernant 1.357 patients et de définir des facteurs de risque à prendre en compte dans la prise en charge clinique de ces cancers de mauvais pronostic, en combinant le séquençage de l'ADN et le séquençage de l'ARNm (protocole BIG-1, randomisations R1 et R2, NCT02416388 <https://clinicaltrials.gov/study/NCT02416388>). Le séquençage de l'ADN permet d'identifier les mutations ponctuelles, les petites insertions et les délétions. Le séquençage de l'ARN identifie également les mutations ponctuelles, les petites insertions et les délétions mais aussi les gènes de fusion. Il permet également de générer le phénotype moléculaire défini par le transcriptome des cellules blastiques médullaires au moment du diagnostic.

La combinaison des données cliniques et génomiques au diagnostic ainsi que l'évolution clinique des patients permet d'étudier les propriétés spectrales et structurelles de la matrice obtenue pour détecter la structure des blocs cachés de manière non supervisée. Ces motifs extraits permettront de caractériser des classes de patients aux profils génomiques similaires. Ces méthodes seront basées sur la factorisation de matrice non négatives et des techniques d'apprentissage non supervisé. Ensuite un modèle prédictif non linéaire basé sur XgBoost et/ou réseaux de neurones sera conçu afin de prédire la survie des patients.

L'ensemble des données sont immédiatement disponibles.

## 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: [genopheno@stages-medecine-numerique.fr](mailto:genopheno@stages-medecine-numerique.fr)
- > Date limite de candidature : 31/01/2024
- > Encadrants: Professeur Eric Delabesse (CHU de Toulouse/IUCT-Oncopole/CRCT), Sandrine Mouysset (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>.