

## PROPOSITION DE STAGE de M1 (2021)

Analyse de la signature mutationnelle d'une nitrosamine dérivée du tabac dans les données de génomes de cancers humains

### Lieu :

Centre International de Recherche sur le Cancer (CIRC/IARC)

Branche Epigénomique et mécanismes (EGM)

[Anciennement groupe mécanismes moléculaires et mécanismes (MMB)]

150 Cours Albert Thomas, F-69372 Lyon CEDEX 08, France

### Présentation de la société et du secteur d'activité

Le CIRC fait partie de l'Organisation Mondiale de la Santé (OMS), institution spécialisée des Nations Unies. La mission du CIRC consiste à coordonner et à mener des recherches sur les causes du cancer chez l'homme et sur les mécanismes de la cancérogenèse, ainsi qu'à élaborer des stratégies scientifiques de lutte contre le cancer. Le Centre participe à des recherches épidémiologiques et expérimentales, et assure la diffusion de l'information scientifique au moyen de publications, de bases de données, de conférences et de cours.

Les objectifs et publications du groupe sont présentés ici :

<https://iarc.who.int/research-groups-mca-mmb-researchtopics/>

### Responsables :

Vice-responsable de la branche : Dr Jiri Zavadil

Chef de projet : Dr Michael Korenjak

Responsable (contact) : Dr Claire Renard (RenardC@iarc.fr)

### Description du stage :

Les signatures mutationnelles sont des signaux mathématiques spécifiques à certains types de mutations et sont utilisées comme empreintes un processus de mutation spécifique, comme par exemple l'exposition à des réactifs causant des dommages à l'ADN. Des signatures mutationnelles ont été extraites à partir de milliers de génomes de cancers humains et sont regroupées dans le « Catalogue of Somatic Mutations in Cancer (COSMIC) ». Elles peuvent être basées sur le changement d'un nucléotide (Single-base substitution, SBS), d'une paire de nucléotide (Doublet-Base Substitution, DBS), sur des insertions-délétions spécifiques (InDel) ou des variations structurales à plus grande échelle.

Des études épidémiologiques ont trouvé un lien entre la consommation de tabac et au moins 12 types de cancer. La fumée de cigarette contient plus de 7000 composés chimiques, parmi lesquels plus de 70 sont responsables d'induction de cancer dans des modèles en laboratoire. Dans le cadre d'exposition contrôlée de lignées cellulaires humaines, suivi de séquençage d'exomes (WES) ou de génomes entiers (WGS), notre groupe a identifié une signature mutationnelle potentiellement nouvelle liées à l'exposition d'une nitrosamine spécifique du tabac.

Le but de ce stage sera de passer au crible les données publiques de séquençage de génomes de cancer pour chercher la présence de cette nouvelle signature. Notamment, le candidat devra :

- 1) Prendre en main et tester un pipeline de screening écrit en python et nextflow, développé en interne, basée sur l'algorithme NNLS (non-negative least square, moindres carrés non-négatifs)
- 2) Appliquer ce pipeline aux données de séquençage de génomes des cancers de la base données PCAWG (Pan-Cancer Analysis of Whole Genome), notamment pour les cancers liés au tabac, pour chercher la présence éventuelle de cette signature spécifique.

Le stagiaire aura un ordinateur dédié, sera encadrée par une bioinformaticienne du groupe (en français ou en anglais, selon préférence) et interagira régulièrement avec les porteurs du projet (Dr. Michael Korenjak, CR, et Dr. Jiri Zavadil, co-responsable de la branche, en anglais), ainsi qu'avec les autres membres de la branche et les autres bioinformaticiens et biostatisticiens de l'institut.

**Mots clés** : cancers liés au tabac, signature mutationnelle, algorithme NNLS

**Compétences requises/Profil de recherche** : Étudiant.e enthousiaste et motivé.e possédant des compétences en programmation (python ou R) et à l'aise dans un environnement UNIX, une expérience du travail sur un cluster et la connaissance du langage nextflow, et / ou la gestion de données de séquençage WGS serait un plus. Une bonne maîtrise de l'anglais est nécessaire pour communiquer avec les porteurs du projet.

**Indemnité** :

- 700 Euros/mois

**Durée** :

3-4 mois

### M1 INTERNSHIP PROPOSAL (2021)

Analysis of a tobacco-specific nitrosamine-derived mutational signature in human cancer genome data

**Place:**

International Agency for Research on Cancer (CIRC/IARC)  
Epigenomics and Mechanisms (EGM) Branch  
[ex. Molecular Mechanism and Biomarker (MMB) group]  
150 Cours Albert Thomas, F-69372 Lyon CEDEX 08, France

The International Agency for Research on Cancer (IARC) is the specialized cancer agency of the World Health Organization. The mission of IARC is to promote international collaborations in

cancer research to identify the causes of cancer so that preventive measures may be adopted and the burden of disease and associated suffering reduced. IARC is a highly inter-disciplinary and collaborative environment, combining approaches in epidemiology, biostatistics, laboratory sciences, bioinformatics and biobanking.

To know more about the group specific objectives and publications:

<https://iarc.who.int/research-groups-mca-mmb-researchtopics/>

### **Supervisors:**

Deputy branch head: Dr Jiri Zavadil

Project leader: Dr Michael Korenjak

Supervisor (contact): Dr Claire Renard ([RenardC@iarc.fr](mailto:RenardC@iarc.fr))

### **Internship description:**

Mutational signatures, specific mathematical readouts of global mutagenic effects, are distinctive patterns (or footprints) caused by specific mutagenic processes such as exposure to individual DNA damaging chemicals [1]. These mutational signatures extracted from thousands of cancer genomes have been compiled in the Catalogue of Somatic Mutations in Cancer (COSMIC) [2]. They are based on single-base substitutions (SBS), doublet-base substitutions (DBS), small insertions and deletions (indels), and large-scale genomic structural alterations.

Epidemiological studies have linked tobacco use to at least 12 types of cancer. Tobacco smoke contains >7000 chemicals, of which  $\geq 70$  cause cancer in laboratory models. Using controlled exposure experiments in human cell lines, followed by whole-exome and whole-genome sequencing, our group has identified a characteristic mutational signature associated with exposure to a tobacco-specific nitrosamine.

The purpose of the internship will be to screen public cancer genome data for the presence of the identified signature. The objectives of the internship include:

- 1) Implement and test an in-house available bioinformatics screening pipeline, based on the non-negative least squares (NNLS) algorithm.
- 2) Analyze the genomes from tobacco-associated cancers, available from the Pan-Cancer Analysis of Whole Genomes (PCAWG) Consortium, for the presence of the new signature.

The trainee will be provided with his/her personal workstation, will be supervised by the group bioinformatician and will interact closely with members of the MMB group and IARC bioinformaticians and biostatisticians.

**Key words:** tobacco-related cancers, mutational signature, NNLS algorithm

**Required skills/Profile:** Enthusiastic and motivated student with programming skills (python or R). Experience in working on cluster with nextflow language, and / or handling genome-wide sequencing data would be an additional qualification. A good level of English language proficiency is required.

**Allowance:**

- 700 Euros/month

**Duration:**

3-4 months