International Agency for Research on Cancer

World Health Organization

150 cours Albert Thomas
69372 Lyon cedex 08, France

Infections and Cancer Biology Group
Tel.: +33 4 72 73 81 90
Fax: +33 4 72 73 86 63
E-mail: icb@iarc.fr
http://www.iarc.fr

Ref.: ICB/AR   29 October 2019

Master 2 internship bioinformatics research project (2020 spring term)

Section: Section of Infections
Group: Infection and Cancer Biology Group

Supervision: Dr. Massimo Tommasino (tommasinom@iarc.fr)
              Alexis Robitaille (robitaillea@students.iarc.fr)

Title: Detection and characterization of novel *Papillomaviridae* sequences in DNA samples

Summary: Papillomaviruses (PVs) are widely distributed across vertebrates and are classified into genera, species, and types based on the nucleotide sequence identity of the major capsid protein L1 [1, 2]. The detection of known human papillomaviruses (PVs) from targeted wet-lab approaches has traditionally used PCR-based methods coupled with Sanger sequencing. With the introduction of next-generation sequencing (NGS), these approaches can be revisited to integrate the sequencing power of NGS. We have recently developed a tool for the classification and identification of novel viral sequences from data produced by amplicon-based methods (https://github.com/IARCbioinfo/PVAmpliconFinder). This analysis workflow is designed to rapidly identify and classify known and potentially new *Papillomaviridae* sequences from NGS amplicon sequencing with degenerate PV primers [3].

Aims: The specific aims of the project will be:

- o Optimize the algorithm to detect PVs from amplicon-sequencing;
- o Explore new methodologies to improve the detection and taxonomic classification of PVs sequences;
- o Participate in the different projects conducted by the team by providing a support for the data analysis;
- o Collaborate with a multi-disciplinary team, and with other bioinformaticians in the Agency.

Required skills: Under the supervision of Dr. Tommasino and the co-supervision of Alexis Robitaille, you will perform analysis of NGS and genomic data using various different existing tools and to perform statistical tests. It is thus important to have very good knowledge in statistics (R software), in Python/Perl and Shell programming, but also a great interest in molecular evolution. You will have to be

independent and autonomous rapidly to share your expertise with other members of the team.

Context: The International Agency for Research on Cancer (IARC) is the specialized cancer agency of the World Health Organization. The objective of the IARC is to promote international collaboration in cancer research. The Agency is inter-disciplinary, bringing together skills in epidemiology, laboratory sciences and biostatistics to identify the causes of cancer so that preventive measures may be adopted and the burden of disease and associated suffering reduced. A significant feature of the IARC is its expertise in coordinating research across countries and organizations; its independent role as an international organization facilitates this activity. The Agency has a particular interest in conducting research in low and middle-income countries through partnerships and collaborations with researchers in these regions.

References:
1. Bzhalava D, Eklund C, Dillner J. International standardization and classification of human papillomavirus types. Virology. 2015;476:341–4.
2. de Villiers E-M. Cross-roads in the classification of papillomaviruses. Virology. 2013;445:2–10.
3. Brancaccio RN, Robitaille A, Dutta S, Cuenin C, Santare D, Skenders G, et al. Generation of a novel next-generation sequencing-based method for the isolation of new human papillomavirus types. Virology. 2018;520:1–10.

Alexis Robitaille, PhD student
Bioinformatics, Infection and Cancer Biology Group
robitaillea@students.iarc.fr