

# NGS4Cloud: Cloud-based NGS Data Processing

Authors: Alexandre Almeida nº40640, João Forja nº41087

Supervisors: Cátia Vaz (ISEL), José Simão (ISEL), Alexandre P. Francisco (IST)

Instituto Superior de Engenharia de Lisboa

Projeto e Seminário – Semestre de Verão 15/16

## Background

Next-Generation Sequencing (NGS) technologies are greatly increasing the amount of genomic computer data, revolutionizing the biosciences field and leading to the development of more complex NGS Data Analysis techniques. These pipelines involve the use of multiple software tools and data resources in a staged fashion, with the output of one tool being passed as input to the next one.

Due to the complexity of configuring and parametrizing pipelines, the use of NGS Data Analysis techniques is not an easy task for a user without IT knowledge. Moreover, knowing input data can be as much as terabytes and petabytes, the execution of pipelines require, in general, a great amount of computational resources and time.

## Objectives

- Make the design and parametrization of a pipeline user friendly and easy for non-IT experts;
- Facilitate the reproducibility of a pipeline;
- Run the pipelines efficiently, making use of a cluster of powerful machines and parallelization techniques.

## Solution

