H2020 Application

Title: Unveiling the mechanism of Disease X

Acronym: EuroX

Participants:

- Institute 1, Country A [Coordinator]
  - P1 - PI
  - P2 - Bioinformatics
  - P3 - Genomics
- Hospital 1, Country A
  - P4 - Clinical Data
- Institute 2, Country B
  - P5 - Data Management
  - P6 - Ethics
- Hospital 2, Country B
  - P7 - Clinical Data
- Institute 3, Country C
  - P8 - Bioinformatics
  - P9 - Genomics
- ...

Start date: January 1st, 2021

Duration: 36 months

Abstract:

The cause of Disease X has been recently discovered to be a virus, phage X, which infects normal gut bacteria and leads them to become virulent and cause chronic intestinal infection. Although non-fatal, this disease has been spreading rapidly in Europe, with costs in health-care reaching the tens of millions of Euros.
This project aims to uncover the mechanisms of disease X by sequencing phage X and tracking its variants across Europe, as well as by studying the effects of its infection in human gut microbiota at the population and molecular level. We will assess which bacterial taxa are infected by phage X and what effect the infection has on the relative abundance of the various taxa, as well as what effect the infection has on the infected taxa at the gene expression level. Furthermore, we will assess whether there dietary habits have an impact in contracting the disease and on its symptoms.
The project will be a key step towards improving treatment for Disease X, and potentially being able to cure it.

Research plan and method summary:

The project will be divided into four activities:

1. Clinical data and sample collection
2. Phage X sequencing
3. 16S sequencing
4. Metatranscriptomics

In the sample collection activity, we will define a study group of volunteer disease X patients across Europe, numbering no less than 2000, and a control group comprising their healthy close relatives, 1-2 per patient. We will collect stool samples from each of the volunteers, and survey them with respect to dietary habits.

In the Phage X sequencing activity, we will carry out DNA sequencing of the stool samples and assemble the genome of Phage X, identifying its variants across Europe. In order to facilitate the assembly while enabling the reliable identification of sequence variants, we will combine the higher quality but short read sequencing technology of the Illumina NextSeq 500 sequencer with the long read but lower quality technology of the Nanopore MinION sequencer.

In the 16S sequencing activity, we will do microbial diversity analysis of the stool samples based on DNA sequencing of the 16S rRNA gene, to assess the impact of Phage X infection of the diversity and relative abundance of gut bacteria taxa. This sequencing will be carried out in Illumina NextSeq 500 sequencers.

In the metatranscriptomics activity, we will assess the effect of Phage X infection at the gene expression level through RNA sequencing of the samples. This sequencing will be carried out in Illumina NextSeq 500 sequencers.

Expected Data & Metadata Outputs:

1. Sample Collection:
    - Patient clinical data (< 1 GB)
    - Sample identification table (< 1 GB)
2. Phage X sequencing
    - Raw FASTQ sequencing data - NextSeq (60 GB)
    - Sample preparation & sequencing metadata - NextSeq (< 1 GB)
    - Raw FASTQ sequencing data - MinION (1 TB)
    - Sample preparation & sequencing metadata - MinION (< 1 GB)
    - Assembled Phage X genome (< 1 GB)
    - Assembly metadata (< 1 GB)
3. 16S sequencing
    - Raw FASTQ sequencing data (15 TB)
    - Sample preparation & sequencing metadata - NextSeq (< 1 GB)
    - Biome tables (< 1 GB)
4. Metatranscriptomics
    - Raw FASTQ sequencing data (120 TB)
    - Sample preparation & sequencing metadata - NextSeq (< 1 GB)
    - RNAseq count tables (< 1 GB)
    - Differential expression test results (< 1 GB)