

## Course Title: Omic Technologies: Variant Calling Using “NGS” Data

**Modality: CFT- Transversal Training Course**

### Orientation:

- Sustainable use of Marine Resources (Field: Aquaculture; Research Lines: Genetics and genomics applied to aquaculture)
- Integral Management of the Sea (Field: Environmental Analysis and Evaluation; Research Lines: Biodiversity and coastal ecology)

### Dates:

Thursday, 21 November 2019

Friday, 22 November 2019

### Timetable:

10:00-13:00

14:30-16:30

### Duration:

Lectures: 10 hours (plus 15 hours of online work)

### Location:

DOMAR Videoconference Room, Torre-CACTI building, Campus As Lagoas-Marcosende, Universidade de Vigo

**Language:** English

### Academic coordinators:

Name	Institution	e-mail
David Posada	Universidade de Vigo	dposada@uvigo.es

### Lecturers:

Name	Institution	e-mail
Fernando Cruz	Centro Nacional de Análisis Genómico-Centro de Regulación Genómica (CNAG-CRG, Barcelona)	fernando.cruz@cnag.crg.eu

### General description:

This course offers an overview of current massive parallel sequencing data (NGS) paying special attention to those produced by the Illumina sequencing technology and its main applications to Genomics Research. Additionally, provides a basic training in bioinformatics techniques for handling, managing and analysing NGS data. It will follow the mapping-based approach for calling Single Nucleotide Variants (SNVs) recommended by GATK (Broad Institute). Additionally, we will

do several exercises for handling, filtering and extracting information from VCFs and GVCFs. Finally, these data will allow the estimation of basic statistics in Population Genomics.

**Contents:**

1. Classification of Genomic Variants.
2. Massive Parallel Sequencing Data.
3. Standard File Formats for sequencing and genomic variants data.
4. Bioinformatics Tools (FastQC, BWA, Picard, GATK, VCFTools...)
5. Mapping-based approach for Variant Calling (GATK Best Practices)
6. Progress in the field.

**Teaching methodologies:**

The course will focus on the basic tools and approaches for variant calling using NGS data, paying special attention to the SNV class. It will be organized in two classroom sessions of 5 hour each. They will be completed with 15 hours of online exercises that will require the students to work individually. The classes will include short practical exercises to facilitate the assimilation of theoretical concepts and methodological approaches.

**Evaluation system:**

The assistance and participation of the following items:

- Lectures: 20%
- Classroom exercises: 30%
- Online exercises: 50%

**Brief CV of the lecturers:**

Fernando Cruz obtained his PhD at the Department of Genetics (University of Vigo). Then, he worked as postdoctoral researcher in the Smurfit Institute of Genetics (Trinity College Dublin, Ireland), the EBC (Uppsala University, Sweden), the Evolutionary Bioinformatics Laboratory (Department of Ecology and Evolution, University of Lausanne) and the Doñana Biological Station (EBD-CSIC, Seville), gaining considerable knowledge and experience in Bioinformatics, Programming and Genomics. Since February 2014, he has been working as Bioinformatician at the National Centre for Genomic Analysis (CNAG-CRG, Spain). He contributed to 15 peer-reviewed publications on the fields of Bioinformatics, Genomics, Evolutionary Biology and Genome Assembly. Please visit <https://publons.com/researcher/2711489/fernando-cruz/>

**Relevant references:**

These four are related with SNPs and Population Genomics:

1. The Lynx genome paper (2016) - Genome Biology - DOI: 10.1186/s13059-016-1090-1
2. SNP panel in Lynx (2017) - BMC Genomics - DOI: 10.1186/s12864-017-3946-5
3. Dog domestication paper (2008) - MBE - DOI: 10.1093/molbev/msn177
4. Modelling marine snail polymorphism (2005)- JEB - DOI: 10.1111/j.1420-9101.2004.00773.x

**Teaching experience includes:**

Lecturer in the Doctorate Course “Variant Calling Using NGS Data” (Campus Do Mar, University of Vigo). Teaching time: 25h. Two seasons: October 2015 and 2017. Campus Do Mar, Vigo (Spain). Assistant during the practical sessions of the Course “Introduction to the Analysis of Next Generation Sequencing Data” taught by Matthew T. Webster (Uppsala University). Teaching time: 19h. 5-8 March 2013, EBD-CSIC, Seville (Spain).