

Erika Ferrandi

CURRICULUM VITAE ET STUDIORUM

Personal information

ORCID ID: 0000-0002-2968-8350

Education and training

October 2018 – December 2020

Qualification: Master Degree in Molecular Biotechnology and Bioinformatics
Institution: Università degli Studi di Milano (Facoltà di Scienze e Tecnologie)
Final grade: 110/110 *cum laude*
Thesis: Bioinformatics approaches for the study of Allele Specific Expression (ASE) in single individuals

September 2015 – July 2018

Qualification: Bachelor degree in Biotechnologie Mediche
Institution: Università degli Studi di Milano (Facoltà di Medicina e Chirurgia)
Final grade: 105/110
Thesis: A bioinformatics approach to study the role of transposable elements in the hematopoietic system

September 2010 – July 2015

Qualification: Secondary School Diploma
Institution: Liceo Scientifico e Linguistico Statale Paolo Giovio (Como)
Final grade: 100/100

Professional experiences in the scientific field

June 1st 2023 – May 31st 2024

Research Grant (Type A – “Professionalising Grants”)

Position: Research fellow
Institution: Istituto di Biomembrane, Bioenergetica e Biotecnologie Molecolari - Consiglio Nazionale delle Ricerche (CNR) of Bari in agreement with Dipartimento di Bioscienze of Università degli Studi di Milano (UNIMI)
Supervisor: Prof. Graziano Pesole
Tutor: Prof. Matteo Chiara
Project: 1. Development of bioinformatics resources for the management and analysis of human data in the context of the Horizon Europe project Digital Europe Programme - European Genomic Data Infrastructure (GDI)

April 1st 2021 – March 31st 2023

Research Grant (Type A – “Professionalising Grants”)

Position: Research fellow
Institution: Istituto di Biomembrane, Bioenergetica e Biotecnologie Molecolari - Consiglio Nazionale delle Ricerche (CNR) of Bari in agreement with Dipartimento di Bioscienze of Università degli Studi di Milano (UNIMI)
Supervisor: Prof. Graziano Pesole
Tutor: Prof. Matteo Chiara
Project: 1. Development of methods and algorithms for tracking SARS-CoV-2 variants and for their functional annotation
2. Scientific coordination of the COVID-19 Data Portal Italy

October 2019 – November 2020

Curricular internship

Position: Master degree graduate student
Institution: Università degli Studi di Milano. Dipartimento di Bioscienze
Supervisor: Prof. Matteo Chiara
Objectives: 1. Implementation and validation of bioinformatics workflows for the identification of genetic variants in the human genome
2. Development and testing of bioinformatics methods for the analysis of Allele Specific Expression (ASE) of genes

February 2018 – June 2018

Curricular internship

Position: Bachelor degree graduate student
Institution: Fondazione INGM. Bodega Lab
Supervisor: Beatrice Bodega
Tutor: Shruti Sinha
Objectives: 1. Bioinformatics analysis of patterns of expression of repetitive genomic elements (especially LINE) from bulk RNA sequencing data

Research activity

During both curricular internships I took part in research projects in which I developed and applied bioinformatics methods for the analysis of high throughput sequencing data, obtained by means of NGS (Next Generation Sequencing) technologies. During these formative experiences, I was able to acquire a good knowledge of basic skills required to correctly manage and analyse large amounts of data. More in detail, I was able to learn how to deal with both gene expression and genome sequencing data for the analysis of genetic variants. Moreover, since I contributed to the development and testing of novel bioinformatics methods, I matured an advanced knowledge of the Python and R programming languages and a good level of autonomy in using Linux/Unix systems. I developed the majority of the workflows used during my master thesis project using the CWL language, a renowned international open standard for workflow description that makes bioinformatics analyses easily reproducible and applicable in different contexts and workspaces. These aspects are particularly important as they ensure results reproducibility and data availability according to the principles of “open science” and “FAIRness”.

A selection of the workflows and scripts I developed during my master thesis project is available at this [link](#).

Starting from April 2021 until March 2023, I worked as a research fellow (Type A - “Professionalizing Grants”) for Istituto di Biomembrane, Bioenergetica e Biotecnologie Molecolari - Centro Nazionale delle Ricerche (CNR) of Bari within the Elixir IIB project. Throughout the whole duration of the research grant, my activity was carried out at Dipartimento di Bioscienze of Università degli Studi di Milano (UNIMI).

My research activity during this period revolved around two main tasks:

I. Development of methods and algorithms for tracking SARS-CoV-2 genetic variants and for their functional annotation

This objective was achieved by performing the projects listed below:

- Weekly analysis of the genomic sequences of SARS-CoV-2 collected from publicly accessible repositories. Data was analysed using custom tools and workflows.
- Generation of informative reports and graphical representations of SARS-CoV-2 variants data. These analyses were initially produced in collaboration with Istituto Superiore di Sanità (ISS) with the purpose of monitoring the circulation of epidemiologically relevant SARS-CoV-2 variants (Variants of Concern or VOCs) in Italy.
- Development of a standalone application for monitoring the circulation of SARS-CoV-2 variants, lineages and mutations at geographic level. This app was developed using R Shiny and consists of a sophisticated interactive user interface, which allows users to select and visualise data in real time. For example, users can select a specific country and/or a specific interval of time, and obtain several informative graphics that provide information concerning the weekly prevalence of a collection of variants, lineages or mutations, their regional distribution, the number of SARS-CoV-2 genomes sequenced every week and many others. The code for the application is available at the following [link](#).

II. Scientific coordination of the COVID-19 Data Portal Italy

The COVID-19 Data Portal Italy is a platform that provides information, guidelines, tools and services to support researchers in producing and sharing with the scientific community data and results about SARS-CoV-2 and COVID-19. The Italian Portal is managed by [ELIXIR-IT](#), in collaboration with CNR, GARR and ISS, as part of a Federative European Initiative promoted by [EMBL-EBI](#). The COVID-19 Data Portal Italy was shut down during May 2024, following the end of the research project funding it. A complete backup of its contents is available through the following [link](#).

As the Content Manager, I had to deal with different aspects related to the management of the COVID-19 Data Portal Italy (Portal from now on) and its contents. More in detail:

- Production of informative scientific material (Highlights on the Portal) on SARS-CoV-2 and COVID-19. I produced and published (bilingual version, Italian + English) 38 Highlights in total.
- Creation and management of the Dissemination & Training page of the Portal (bilingual version, Italian + English). This resource is dedicated to the promotion of events, seminars and workshops on SARS-CoV-2, COVID-19 and the Open Data and Open Science principles in data sharing.
- Periodical evaluation and approval of contents reported by Portal users and direct interaction with the users. To maximise its outreach the Portal hosts a section of user provided content, which serves as a showcase of ongoing research activities and projects on SARS-CoV-2 and COVID-19 in Italy. Prior to publication on the Portal, each content is carefully evaluated according to a series of guidelines. Once approved these contents are sorted by thematic categories and published on the dedicated page of the Portal.
- Communication with the users through the activation and management of a Help Desk.
- From August 2021 to March 2023, curation of a Newsletter focused on the promotion of new contents from the Portal and on high impact scientific papers (from both Italian and International authors) about SARS-CoV-2 and COVID-19 research effort.

Starting from June 2023 until May 2024, I worked as a research fellow (Type A - “Professionalizing Grants”) for Istituto di Biomembrane, Bioenergetica e Biotecnologie Molecolari - Centro Nazionale delle Ricerche (CNR) of Bari within the Horizon Europe Digital Europe Programme - European Genomic Data Infrastructure (GDI). Throughout the whole duration of the research grant, my activity was carried out at Dipartimento di Bioscienze of Università degli Studi di Milano (UNIMI).

My research activity during this period revolved around two main tasks, together with a third collateral task:

I. Taking part to the “Pillar II” activities of the “Genomic Data Infrastructure” (GDI) project.

GDI is a project funded by the European Commission, which main purpose is the creation of a sustainable, reliable and federated infrastructure that allows to access European genomic, phenotyping and clinical data in a simple and swift way. More in detail, the “Pillar II” work group guarantees that data shared and examined using this infrastructure respect the FAIRness (Findable, Accessible, Interoperable, Reusable) principles of data sharing. This task requires the development of tools and national services based on community guidelines which define the minimal requirements for the correct deployment of the described infrastructure. The GDI “Starter Kit”, a

collection of tools that allows the implementation of national nodes according to community standards, was developed taking in account the abovementioned guidelines and minimal requirements. All the tools and methods proposed were tested and validated by analysing a collection of “use cases”, including:

- Rare diseases
- Cancer
- Infectious diseases
- Complex diseases
- Population genomics

In this context my main activities within GDI consisted of:

- Taking part to 9 meetings in total (from June 2023 to May 2024, monthly basis). The purpose of these meetings was discussing future tasks and resolve technical issues. For the same reasons I also took part to the GDI First Annual Meeting in Rome in October 2023.
- Installing and testing “Beacon”, a tool from the “Starter Kit” for aggregated genomic data sharing and searching.
- Development of tools and methods for the visualisation and sharing of genomic data from human pathogens, in accord with the GDI task regarding the development of use cases about infectious diseases.

II. “Beacon” usage and evaluation.

“Beacon” is a protocol developed by the “Global Alliance for Genomics and Health” (GA4GH) aiming to facilitate the collection, search and access to genomic and biomedical data into federated databases in order to help research and clinical practice. The current iteration of “Beacon”, called “Beacon v2”, is included in the GDI “Starter Kit”.

In these regards, my main goal was testing “Beacon v2” functionality and application as an end-user, stressing on the possibility of applying it to research project in the context of the Italiana node of GDI.

III. Development and testing of the “mapPat” application

While working on GDI related projects I completed the development of “mapPat”, an application for monitoring the spreading of pathogens during epidemics and/or pandemics. More in detail, “mapPat” allows to monitor the spreading of pathogens’ variants, lineages and mutations (at the moment, only SARS-CoV-2 and mPox data are available) in time and space. The application consists of:

- An interactive user interface collecting a series of widgets for the personalization of data visualization and filtering.
- Three tabs (representing variants, lineages and mutations respectively) each consisting of a different collection of plots describing data distribution at different levels of granularity (from simple area charts to more complex choropleth maps). Graphical representations are produced based on the data selection performed by the user.

The complete code for “mapPat” is available through the [dedicated GitHub repository](#), together with a workflow for generating all the inputs required by the application.

Participation to international projects

The scientific activities outlined above were carried out in the context of internationally financed scientific projects, in particular:

- **ELIXIR-CONVERGE Work Package 9 (WP9) - Grant agreement ID: 871075**
Project that promotes open sharing of data and methods for research on SARS-CoV-2 and COVID-19 through open source platforms. By doing so information on the novel Coronavirus will become more easily accessible for analysis and used to monitor the appearance and/or evolution of new SARS-CoV-2 variants. Further information available at the following [link](#).
- **BeYond-COVID (BY-COVID) - Grant agreement ID: 101046203**
Project that promotes the development of resources, shared models and guidelines that effectively respond to all the issues derived from the huge amount of data produced during a pandemic. Further information available at the following [link](#).
- **European Genomic Data Infrastructure (GDI) – Grant agreement ID: 101081813**
Project that promotes the development of a European federated infrastructure enabling secure and controlled sharing and access to human genetic data and related clinical data. Further information available at the following [link](#).

Participation to international conferences

- **Galaxy Community Conference 2021 (GCC2021) – July 6th-8th - Virtual edition**
Participation as a speaker with a talk named “**CorGAT and CorGAT-tracker: functional annotation of SARS-CoV-2 genome and tracking mutations and variants of concern**”. A recording of the speech and the respective slides can be visualised at the following [link](#).
- **GDI First Annual Meeting – October 9th-10th-11th 2023 – Rome**
First annual meeting of the GDI project. During the three days span of the meeting representatives from all GDI contributors were invited to share the advancement state of each node and to discuss future perspectives for the following years of the project. Moreover, the different workshops available allowed participants to compare their experiences and establish profitable collaborations.

Published papers

- Zambelli F, Chiara M, **Ferrandi E**, Mandreoli P, Tangaro MA, Pavesi G, Pesole G. **aScan: A Novel Method for the Study of Allele Specific Expression in Single Individuals**. J Mol Biol. 2021 Jan 26. doi:[10.1016/j.jmb.2021.166829](https://doi.org/10.1016/j.jmb.2021.166829).

- Chiara M, Horner DS, **Ferrandi E**, Gissi C, Pesole G.
HaploCoV: unsupervised classification and rapid detection of novel emerging variants of SARS-CoV-2. Communications Biology. 2023 Apr 22. doi:[10.1038/s42003-023-04784-4](https://doi.org/10.1038/s42003-023-04784-4).
- Pinoli P, Canakoglu A, Ceri S, Chiara M, **Ferrandi E**, Minotti L, Bernasconi A.
VariantHunter: a method and tool for fast detection of emerging SARS-CoV-2 variants. Database. 2023 Jul 06. doi:[10.1093/database/baad044](https://doi.org/10.1093/database/baad044).