

# Elena Buscaroli



---

## Education and training

---

### PhD in Applied Data Science and Artificial Intelligence

November 2022 - expected February 2025, Cancer Data Science Lab, Università degli Studi di Trieste  
Currently enrolled in the third year in the Life Sciences and Medicine curriculum.

Thesis: *Bayesian inference of lineage evolution and molecular profiling in cancer and gene therapy.*  
Supervisor: Prof. Giulio Caravagna

Main topics: Development of Bayesian models and related softwares for genomic data analysis.

### Master degree in Data Science and Scientific Computing (Class LM-44)

September 2020 - October 2022, Università degli Studi di Trieste and SISSA

Thesis: *Bayesian multi-lineage tracing in gene therapy assays.*

Supervisor: Prof. Giulio Caravagna

Final mark: 110/110 cum Laude

Main topics: Statistical modeling and machine learning, data science, high performance computing, mathematical optimization, deep learning.

### Bachelor degree in Genomics (Class L-2)

September 2017 - September 2020, Università degli studi di Bologna

Thesis: *Accurate identification of low allelic frequency mosaicism in targeted sequencing experiments.*

Supervisor: Prof. Tommaso Pippucci

Final mark: 110/110 cum Laude

Main topics: Molecular biology, genetics, bioinformatics, programming, statistical modeling, data science.

---

## Research experience

---

### PhD visiting student

June 2025 - August 2025, Statistical Intelligence for Cell Biomedicine Lab, Hong Kong University

Supervisor: Prof. Yuanhua Huang

Project: *Modelling of clonal somatic variations from multi-modal single-cell and spatial omics data.*

Development of latent Gaussian process models to improve detection and denoising of somatic mutation abundances, leveraging phylogenetic or spatial covariance among cells in structured single-cell populations.

### Research Intern - Master degree thesis project

November 2021 - September 2022, Cancer Data Science Lab, University of Trieste

Supervisor: Prof. Giulio Caravagna

Project: *Bayesian multi-lineage tracing in gene therapy assays.*

Development of Bayesian models for lineage inference in gene therapy assays, accounting for somatic mutations accumulation, collected in [lineaGT](#) and [pyLineaGT](#) R and Python packages. Project carried out in partnership with San Raffaele Telethon Institute for Gene Therapy.

## Research Intern - Bachelor degree thesis project

March 2020 - June 2020, University of Bologna

Supervisor: Prof. Tommaso Pippucci

Project: *Accurate identification of low allelic frequency mosaicism in targeted sequencing experiments.* Implementations of machine learning algorithms to accurately discriminate sequencing errors from true rare somatic variants in patients affected by epilepsy.

---

## Training experience and certificates

---

### Courses, summer schools and conferences

Working Group on Model-Based Clustering Summer Session. *July 2024, Bertinoro*

MLAS Summer School, Gaussian processes and Bayesian optimization. *June 2024, Madrid*

19th Annual Meeting of the Bioinformatics Italian Society. *June 2023, Bari*

Como school on cancer evolution. *May 2023, Como*

Courses@CRG: Containers and Nextflow. *June 2022, Online*

### Teaching Assistant activities

Computer Science course. Mathematics BSc, University of Trieste.  
*March - June 2022*

Programming course. Artificial Intelligence and Data Analytics BSc, University of Trieste.  
*October - December 2021, 2022, 2023, 2024*

### English certificates

TOEFL certificate, Final score: 95/120. *June 2020*

---

## Skills

---

### Job-related skills

Experience with bioinformatics databases and data banks, and with sequencing data analysis.

Software development methods.

Experience with HPC facilities.

### Digital skills

Python, R, C, C++, Java programming languages. Pyro and Stan PPL, PyTorch framework.

Python and R packages: numpy, scikit-learn, scipy, pandas, OpenMP, MPI, dplyr, ggplot.

Experience with CLI in a Linux-based environment and shell scripting.

Experience with Docker and Singularity and their applications in Nextflow pipelines.

---

## Publications

Calabria, A., Spinozzi, G., Cesana, D., *et al.* **Long-term lineage commitment in haematopoietic stem cell gene therapy.** *Nature* 636, 162–171 (2024). <https://doi.org/10.1038/s41586-024-08250-x>

Buscaroli, E., Sadr, A., *et al.* **BASCULE: Bayesian inference and clustering of mutational signatures leveraging biological priors.** *bioRxiv* (2025); accepted for publication in *Genome Biology*. <https://doi.org/10.1101/2024.09.16.613266>

Rivaroli, E., Buscaroli, E., *et al.* **Model-based tumour subclonal deconvolution accounting for spatio-temporal sampling biases.** *bioRxiv* (2025); accepted in *RECOMB-CCB 2025 Proceedings*. <https://doi.org/10.1101/2025.03.31.646463>

