Teaching Statement

The intersection of high-throughput omics technologies and computer power promises to change basic science and clinical practice in fundamental ways. To keep pace with the outpouring of omics data, we must empower the next generation of scientists with skills that facilitate translating big data into applicable knowledge. My main teaching goal is to equip students with methodological training and hands-on practical experience required to bridge the gap between data and insight.

I achieve this aim by complementing lectures, especially at masters and graduate levels, with practical assignments that close the *knowing-doing* gap. For example, I designed and supervised projects for a bioengineering course (Systems Biology and Bioengineering II: Network Reconstruction), in which participants apply data mining analytics to extract patterns and knowledge from large-scale omics data. One project asks students to analyze genetic information from over 50 strains of *Escherichia coli* by passing data through the stages of alignment, quality control, and mapping to protein structure and networks. At this scale, students run algorithmic workflows through Jupyter notebook frameworks, which are easily implemented on commodity hardware, such as personal laptops. A subsequent exercise asks them to mine patterns in sequence variation to gain insight on strain-specific adaptations across different nutritional environments. In this way, students carry a project all the way from the manipulation of raw sequence data to formally grounded concepts that generate hypothesis-driven research and, ultimately, knowledge. This model for teaching has proven to be productive, as demonstrated by publication of an undergraduate student's work in *BMC Systems Biology*.

Similarly, I designed interactive coding exercises for a graduate-level chemistry course (Molecular Dynamics and Monte Carlo Simulations, EPFL), which teaches the core principles of computational chemistry and data visualization. In accompanying exercises, students model enzyme dynamics, enzyme-ligand interactions, and macromolecular complexes in explicit water models, and apply their newly acquired skills to problems related to drug design. In this way, translation to clinical application motivates further study and links lecture materials to recent discoveries in clinical research.

I ask mentees, at all stages, to formulate clear trajectories for analyzing big data, starting with a data-driven hypothesis that can be formally grounded by published research. I stress the importance of experimental validation to translate data analysis into meaningful insight and I provide hands-on engineering experience rooted in academia and industry. In this respect, I teach standard operating procedures for data analytics, data reproducibility and statistical significance to ensure students practice well-rounded science and are strong competitors on the job market. Recently, I supervised four graduate students and organized their projects around the overarching aim of gaining insight into biological processes through vertical integration of multiple, disparate omics data types. Under my supervision, each student published at least one first author publication, with some publishing as many as three first author publications.

My teaching empowers the next generation of scientists with the tools and concepts to manage and dissect the data they generate. Being able to tackle the proliferation of big biological data will be an asset to any biologist, physician scientist or computer scientist, regardless of career path or academic trajectory.