# Biomedical Big Data Training Grant

## University of California Los Angeles

PI: PELLEGRINI, MATTEO                                        Grant Number: 1T32 CA201160-01

This broad-based T32 proposal is focused on supporting graduate students pursuing Biomedical Big Data Analysis research at UCLA. Over the past few years there has been increasing recognition that the biomedical sciences are undergoing a transformation, led by the development of new technologies that have enormously increased the capacity to generate data. These include technologies to sequence DNA and RNA; measure protein and metabolite abundances using mass spectrometry; as well as multiple other high throughput platforms for screening and phenotyping. Coupled with the advances in medical imaging and EHRs, the amount of data is growing faster than ever before. While several other training grants exist at UCLA, none are specifically focused on the general problem of analyzing and relating large biomedical data sets. Thus this new training grant we propose here fills a critical niche that will allow us to support graduate students in fundamental aspects of biomedical "big data" analysis. This effort realizes the novel development of a tailored set of courses in big data analysis, along with specialized team building activities in bi data challenges and extramural internships in big data centers. This training program will position our students for the future growth in big data science, fostering the growth of this critial area at UCLA. We posit that this program will be critical for the growth of our bioinformatics program and big data biomedical science at UCLA. Faculty supporting this training program represent an interdisciplinary group of researchers from across the UCLA campus, including the Schools of Engineering & Applied Sciences; College of Life Sciences and the Medical School. Many of the faculty are nationally recognized leaders in their disciplines; collectively, these individuals provide a comprehensive and complementary set of (funded) research areas and skills that enrich the training experience. PUBLIC HEALTH RELEVANCE: Since 2008, the UCLA Bioinformatics Program has recruited a growing number of outstanding students trained at the boundary of biology and computational sciences. The proposed Biomedical Big data Training program will support a subset of these students with an interest in developing and applying skills to analyze massive scale biomedical data, such as sequence, proteomics, and medical records. The students will take specialized big data courses and work with pairs of mentors with expertise in big data and either experimental of computational biology.