January 30, 2015

GATP Selection Committee
UCLA Department of Human Genetics
6506 Gonda (Goldschmied) Research Center
Los Angeles, CA 90095-7088

Dear GATP Selection Committee,

First, I would like to thank you all for supporting me with the GATP Training Grant for this past year. It has already made a strong impact in my career at UCLA, most notably by allowing me to make professional connections with members of GATP and to travel to conferences and meet the leaders of our field.

Since the end of the previous academic year, I have been working closely with Dr. Jason Ernst in the field of regulatory genomics, specifically through the lens of machine learning. It has been quite a challenging endeavor so far – we have recently developed a model to predict clusters of regulatory proteins from ChIP Seq data, and using those clusters, to decompose ChIP Seq signal into the sum of its direct components (protein binding to the DNA itself) and indirect components (where the signal is derived either from a long range looping interaction or a protein binding to a complex rather than the DNA itself). Currently, I am in the process of implementing the basic model and writing the code to decompose the signal. It seems promising, and I am about to get my first batch of results.

However, there will be much more work needed to be done to develop it into a robust, complete form, and additional types of data and methods will need to be implemented. More specifically, I am planning on incorporating long-range DNA interaction data (Hi-C data), the general openness or permissiveness of chromatin, and possibly local spatial interactions between clusters. As these assays are relatively new, integration of them is difficult. We are hopeful that it will be the first method of its kind in the field, and that it will be a high impact one. Not only would it be novel, but it would also be quite useful for biologists, as they would be able to run the method (either learning from their own assays or using pre-computed probability matrices) to distill their own data into meaningful and informative results.

In addition to the above project, I will be starting a new project this year. This new project will be a large collaboration between about a dozen labs at UCLA. This collaboration, nicknamed Ribonomics, will study the complex splicing mechanics of RNA. Working with Dr. Ernst, my role will likely include the complex task of deciphering and predicting the unique splicing patterns of a cell, based on exon inclusion levels, genetic markers and possibly even epigenetic markers (I may be able to include predictions from my other project). Incidentally, before I entered the Bioinformatics PhD program, I worked with Dr. Matteo Pellegrini on studying RNA and gene expression levels. I have become quite comfortable with a great number of methods in studying

RNA, and have even looked at splicing prediction. I am looking forward to being able to explore and discover new methods in this subfield.

Over the past half a year, I have also taken classes for both the GATP program and for my own development, as well as attended a series of workshops to work on my IDP. Last quarter, I took a bioinformatics seminar, an ethics course, an online Algorithms course, and a Machine Learning (ML) course, which, incidentally, was the hardest, but most mind opening, class I've ever taken. This quarter, I am taking two seminars (one for bioinformatics, the other for GATP) and a Probabilistic Graphical Models (PGM) course. Together, the algorithms, ML and PGM courses cover a majority of the methods utilized in the field. Nevertheless, I do plan on continuing to take courses to enrich my education and understanding of the genetics field, and I look forward being able to immerse myself in my lab projects.

Thank you very much for your continued support.

Sincerely,

Artur Jaroszewicz
Ph.D. student