Summary of Goals

In a concise statement, provide a summary of your current educational program objectives and your long-range professional goals.

It is crucial that you address each of the following:

- **Why you want to work for the Department of Defense?**
- **How do you think your employment will contribute to the DoD Science and Technology Enterprise?**
- **How do you think that you will benefit from working as part of the DoD civilian workforce?**

(min 2500 chars, max 5000 chars)

Fascinated by my middle school genetics class as the Human Genome Project reached its fruition, I learned early on that the genome is a powerful tool with profound application to understanding disease and became obsessed with genetics. I worked in a nematode lab at a nearby university, where I realized genetics is far more complex than Punnett squares. I attended MIT where I was exposed to high-throughput technologies that inspired me to pursue bioinformatics. In undergrad, I had strong computational biology experiences throughout college, which led to graduating with dual degrees in mathematics and biological engineering, one of two people in class of one thousand.

After graduating, I trained as a bioinformatician in cancer genomics at the Broad Institute of Harvard and MIT. The interdisciplinary environment was inspiring because the algorithm design process was not only an iterative process, but also a convergent symbiotic evolution of biologists and mathematicians. I twice presented my work from the Broad Institute this summer at Intelligent Systems for Molecular Biology (ISMB), the largest bioinformatics conference and at the ISMB Student Council Symposium (SCS), and was invited to co-chair ISMB SCS 2012. Planning this event has been an exciting international experience, collaborating with peers from India, Brazil, Belgium, and Nigeria to continue the circulation of innovations in bioinformatics. I am looking forward to cross-cultural collaboration and providing the opportunity to attend ISMB to students from developing countries.

As an immigrant from the former Soviet Union, I am well aware of the difficulties faced by immigrants and subtle discrimination present in both professional and private life. I sought to ensure that new immigrants had equal opportunity to healthcare through equal access to healthcare and obtained a certificate in Russian-English medical interpretation. Through interpreter training, I was exposed to a variety of ambiguous ethical situations, such as maintaining confidentiality in interpreting for the same patient but for different

practitioners, ameliorating reaction to hearing a cancer diagnosis, and experiencing disrespect to female interpreters—interpreters in Russia are primarily male. I am well aware of the Russian discrimination against women, which my mother was exposed to when she was attending the prestigious Moscow Institute of Physics and Technology, known as the "Russian MIT." There, one of her male peers asked her, "What are you doing here? Why aren't you at home having babies?" To advocate for women in Russia, this summer I will travel to Moscow to help develop the biomedical research programs at the MIT/Skolkovo Institute of Technology (SkTech), where I will organize women's leadership workshops in addition to interviewing potential researchers for SkTech.

My brother Ilia Botvinnik served in the US Marines from 1996-2000, and I have since been interested in contributing to my country. By working as part of the DoD civilian workforce, I will be contributing not only to the body of scientific knowledge, but also to the wellbeing of millions of Americans. Bioinformatics algorithms can be used to precisely target biological warfare and improve soldier infection recovery by pinpointing drug targets and understanding interactions between cells and tissues. Researching computational representations of biology will take on higher meaning as the algorithms I develop will be used to protect both civilian and military citizens. Additionally, as a fluent Russian speaker I have the advantage of collaborating with researchers in Russian-speaking countries and ensuring US technical superiority.

To maximize my impact on revolutionizing bioinformatics, I want to work at the interface of biomedicine and computation as a professor at a research institution or in a national lab, where my quantitative skills will filter noisy biological data into novel discoveries. Seizing opportunities to work with people of diverse experiences and expertise, our strengths together can solve problems in completely unexpected ways. I am now doing a Masters in one year—half the expected time—in Bioinformatics at the University of California, Santa Cruz (UCSC), where I am deepening my understanding of computational biology. By obtaining the DoD SMART fellowship, I would have the freedom to forge high-risk collaborations and work on groundbreaking science.

Interests and Research Explanation

In a concise statement, please elaborate on the kinds of research in which you have engaged and in which you would like to engage during your studies as well as during your expected tenure with the Department of Defense (DoD).

You may wish to discuss specific research questions that interest you and how you became interested in them. Please discuss these research interests in sufficient detail for an expert who is technically competent in your field to judge your understanding of the questions to be addressed, relevant hypotheses and approaches one might take to answering the questions, and other research principles required to investigate in the research area you identify. (min 2500 chars, max 5000 chars)

In my first research experience, I worked in Prof. Martha Bulyk's laboratory where I analyzed homeodomain transcription factor binding sites by performing protein-binding microarray experiments, resulting in a 2008 publication in the journal *Cell*. I was fascinated by how a computer could interpret the minutely polka-dotted microarrays and transform them into binding motifs, inspiring me to further explore quantitative analysis of biology. After my experience in the Bulyk lab, I was admitted to the competitive Janelia Farm (JF) Summer Scholars program to work with Dr. Sean Eddy at Howard Hughes Medical Institute JF Research Campus in Ashburn, Virginia. I used Hidden Markov Models to create a robust null model for HMMER, sequence homology software developed in the Eddy lab. I completed this project in two months in Python and then rewrote it in two months in C, the language of HMMER.

Wishing to continue genome research, I worked in Prof. David Gifford's lab in the MIT Computer Science and Artificial Intelligence Laboratory, where I worked on two projects in yeast metabolic genome networks and T-cell receptor (TCR) sequences. For the yeast metabolome, I hypothesized methods of information content in graphs can inform synthetic lethality of a gene. My metrics indicated purely genomic methods do not fully explain biological interactions. For the TCR project, I created random TCR sequences and tested the minimum frequency a particular sequence must occur for detection by statistical methods. I found that even a slight (2x) increase in frequency was detectable, a promising result for diagnosis. This project was an exciting convergence of bioinformatics and clinical medicine as its implications allow health professionals to screen an individual's vaccination and antigen exposure history.

After graduating, I trained in cancer genomics at the Broad Institute. At Broad, I developed REVEALER, an algorithm that integrates genomic and functional data to infer new associations. A pathway may be broken in many ways, but produce the same phenotypic output. REVEALER finds candidate explanations of a phenotype by removing samples with known causal events and uncovering new events via a mutual information

metric. This ambitious project combines computational methods with genomic analysis in the same way I hope to do in my future research. I twice presented my work from the Broad Institute this summer at Intelligent Systems for Molecular Biology (ISMB) and at the ISMB Student Council Symposium.

Current in vitro models of tumors are wrong. Working at the Broad Institute, I saw results from cell lines extrapolated to a whole class of cancers, but less than 10% of tumor cells even survive in vitro and monoclonal lines do not represent tumor heterogeneity. Certainly, there are successful in vitro cases, such as PLX4032 designed to target BRAF's V600E mutation, but most drugs designed for a cell line are unsuccessful in clinical trials. Before a drug even reaches the clinic, we don't know why applying a drug to monoclonal cell lines results in resistance, and whether survival occurred on a single-cell basis or was driven by intercellular interactions. Relying on statistics and success "most of the time" is an antiquated view – we know have the tools to understand the mechanisms behind an ineffective drug.

I want to shatter current acceptance of monoclonal cell lines as usable models of tumors. I will develop an in silico model of in vitro heterogeneity of breast cancer (BRCA) cell line SK-BR-3 in response to Trastuzumab (Herceptin, a BRCA drug used in HER2 positive cases such as SK-BR-3) by analyzing single-cell RNA-Sequencing (RNA-Seq) data, high-resolution snapshots of a cell's transcriptome and robust determination alternative splice sites. I hypothesize the heterogeneity of a cell population can be modeled by extracting network information from many single-cell transcriptomes. First, I will develop RNA-Seq assemblers tailored to the very small DNA sample amount of single-cell sequencing, less than a nanogram. Second, I will analyze differential transcripts between pairs of neighboring cells, accounting for differences in cell cycle stage using documented cell cycle genes. I expect to see a spectrum of active transcription in secretory pathways of neighboring respondent and healthy survivors, indicating communication between these cells and identifying specific transcripts potentially useful to survival. Using these differentially expressed genes and known interaction networks, I will develop a network of cell-cell interaction between adjacent SK-BR-3 cells in response to Herceptin.

While I use cancer as an example in my proposed research, I will develop methods for use across disease boundaries and throughout biomedicine. This project can be applied to biological warfare by finding mechanisms of resistance and driving development of vaccines and prophylactics against biological agents.

Community and Volunteer Work

Provide a summary of volunteer work and experiences, interests, and/or hobbies. Include the approximate number of hours a week for each experience and a reference for each experience with the reference's name and contact information for verification. If there is no specific contact person, please list someone at the organization who can confirm your experience. (max 400 chars)

We Teach Science Foundation, 8th grade Algebra Tutor (2h/wk 10/11–present) – Camille Stone camille@weteachscience.org

Science Club for Girls, 2nd grade Mentor (4h/wk 1/11-5/11) – Rosalind Gould rgould@scienceclubforgirls.org

Mission Hill Middle School, Math tutor (2h/wk 2/2009–5/2011) – Ann Ruggerio ann.m.ruggiero@gmail.com

Kappa Alpha Theta (2h/wk 9/2007–6/2010) – Victoria Thomas vthomas@mit.edu

Leadership Experiences

Describe or list any leadership experiences, such as class president, committee chairperson, or scouts. Include dates and a reference for each experience, along with the reference's name and contact information for verification. If there is no specific contact person, please list someone at the organization who can confirm your experience. (max 400 chars)

Co-chair, Intelligent Systems for Molecular Biology Student Council Symposium 2012 - Alex Goncearenco neksa@neksa.net
DanceTroupe; Choreographer, Publicity Chair – Catherine Johnson cjohns@mit.edu
Baker House Social Chair (2008) – Jonathan Nolan jnolan@mit.edu
Bernard M. Gordon-MIT Engineering Leadership Program – Leo McGonagle lmcgon@mit.edu
Leadershape – Kirk D. Kolenbrander (617) 253-3365

Teamwork Experiences

Describe or list any teamwork experiences (including research or academic projects as well as varsity sports, extracurricular groups, or clubs). Include dates and a reference for each experience, along with the reference's name and contact information for verification. If there is no specific contact person, please list someone at the organization who can confirm your experience. (max 400 chars)

Co-chair, Intelligent Systems for Molecular Biology Student Council Symposium 2012 - Alex Goncearenco neksa@neksa.net
Biological Engineering Design final project (Spring 2010) – John Essigmann jessig@mit.edu
MIT DanceTroupe; Dancer (Fall 2007 – Spring 2011) – Catherine Johnson cjohns@mit.edu
MIT Lightweight Men's Crew; Coxswain (Fall 2006 – Spring 2007) – Seth Davis fischer.davis@gmail.com