My extensive scientific research experience has exposed me to both the wet lab, empirical approach as well as the dry lab, computational approach of genomic analysis in the study of complex diseases. I plan to apply many of the methods I have developed and skills I have gained through these previous research experiences to accomplish my proposed graduate research.

Delineating the Role of *BRF2* in Breast Cancer Pathogenesis (06/2008 - 08/2008)

In the summer of my junior year in the MBHS Magnet program, I secured a position in the NIH Summer Research Program sponsored by ORWH-NIH-FAES. I interned in the lab of Dr. Paul Meltzer and Dr. Liang Cao to investigate the $8p_{11\text{-}12}$ amplicon, a region of genetic amplification associated with poor breast cancer prognosis. While putative driving oncogenes have been proposed, genes within this amplicon had yet to be definitively implicated in cancer growth, survival, or pathogenesis.

Under the mentorship of Dr. Cao, I evaluated the role of *BRF2*, one gene located within this amplicon, in breast cancer growth and development. I performed immunoblot analysis and used previous CGH and expression array studies to establish a strong correlation between *BRF2* gene amplification and BRF2 protein overexpression, a trait consistent with an oncogenic role. I employed lentiviral-mediated gene transfer to deliver *BRF2*-shRNA into breast cancer cells with $8p_{11\text{-}12}$ amplification and subsequently established long-term stable cell lines with shRNA constructs targeting *BRF2,* leading to marked reduction of BRF2. Using clonogenic assays, I determined that *BRF2* inhibition resulted in impeded growth and proliferation rates as well as increased cell death. My findings suggest that *BRF2* is a relevant oncogene in the $8p_{11\text{-}12}$ amplicon and may play a role in breast cancer growth and pathogenesis. My paper on these findings placed in the semi-finals of the Intel Science Talent Search and Seimens Math, Science, and Technology competitions and was later published in HURJ.

As this was my first research as well as work experience, I learned to trust my own abilities to carry out experiments independently but also ask for help if needed. I learned to allocate my time wisely, assisting others in laboratory tasks while waiting for my gels to run or cells to grow. Although I would soon shift paths from wet lab, empirical analysis to dry lab, computational analysis, this experience opened my eyes to the potential of using DNA and the human genome in the study of complex diseases such as breast cancer.

Computational Assessment of the Utility of Limiting Orthologous Sequence Depth in Mutation Impact Prediction Performance (06/2011 - 05/2012)

Starting my freshman year at Johns Hopkins University, I began interning at the Institute of Computational Medicine under the mentorship of Dr. Rachel Karchin to evaluate the functional impact of mutations using bioinformatics tools and computational models. To predict for the function impact of mutations, current computational models often use sets of orthologous sequences, which are presumed to originate from a common ancestor such that their differences can be attributed to mutation and selective pressures. However, the extent to which these orthologous sequences have been subjected to the same selective pressures and subsequently the validity of using overly deep orthologous sequences remains unknown. In my sophomore year, under the guidance of Dr. Karchin, I wrote a proposal to devise a supervised machine learning approach to assess the utility of limiting orthologous sequence depth in functional impact prediction performance, winning a grant from PURA.

Implementing various R packaged, I trained a soft margin SVM classifier with a general penalizing parameter for C-classification and a radial basis function-specific kernel parameter to optimize on the association of feature scores with the known functional impact of mutation

from an annotated training dataset. I developed the feature scores to capture information concerning the physiochemical differences between reference and variant amino acid residues as well as the evolutionary conservation of amino acid residues up to a certain orthologous sequence depth limit. I measured the overall performance of predictions using standard protocols for statistical learning including calculation of ROC and AUC. My results suggested an orthologous sequence depth limit at the divergence point between vertebrates and invertebrates that may improve mutation impact prediction performance. I presented a poster on these findings at the ICHG/ASHG Conference in Montreal and at various Hopkins research events.

As my first truly independent project, this experience forced me to be critical of my own work. Because I understood the project better than anyone, I was the most equipped to challenge the statistical significance of results or expose any potential underlying confounding errors to ensure that only quality results and methods may be disseminated to the greater community when these research findings submitted to a peer-reviewed journal in the future.

Detecting Synergistic Epistasis in Humans (06/2012 - 08/2012)

Continuing along the path of bioinformatics, this past summer, I seized the opportunity to participate in the Harvard-MIT HST Bioinformatics and Integrative Genomics program. As a part of the program, I interned at Harvard Medical School under the mentorship of Dr. Shamil Sunyaev to devise a theoretical test for detecting synergistic epistasis. The prevalence of sexual reproduction, despite its inherent two-fold cost disadvantage, suggests that sexual reproduction must confer some compensatory evolutionary advantage. The deterministic mutation hypothesis for the evolution of sexual reproduction posits that such an evolutionary advantage may be achieved contingent on synergistic epistasis, whereby accumulations of deleterious mutations lead to larger decreases in relative fitness.

Working with Dr. Sunyaev, we devised a theoretical test to detect for synergistic epistasis in humans using variance-mean ratios of mutations accumulated since the out-of-Africa migration. I applied this test to Genome of the Netherlands (GoNL) data and compared various functional classes of mutations, hypothesizing that variance will be depleted for deleterious mutations but not for benign or neutral mutations due to synergistic epistasis. I distributed analysis over numerous computing nodes on a cluster using parallel processing with the help of instructor Dr. Ivan Adzhubey. I devised and conducted statistical tests including non-parametric bootstrap, ANOVA, and principal component analysis to assess the significance of results. I also performed quality control tests in collaboration with Laurent Francioli of the GoNL consortium to assess for potential batch and flowcell effects. While detection of synergistic epistasis in humans remains inconclusive due to variance inflation from underlying confounding batch and flowcell effects, my results did suggest segregation in variance-mean ratios between benign and damaging mutations, consistent with our hypothesis. I also observed a large spousal correlation in non-coding synonymous mutations, along with no spousal correlation in coding synonymous mutations, suggesting a potential phenotypic readout of mutational load influencing spousal selection, warranting further future investigation. My findings were presented at the BIG Harvard-MIT HST i2b2 Summer Conference.

In working with Dr. Sunyaev, I was inspired by his extensive collaborative network and hope to continue participating in such collaborative efforts on an international scale towards the common goal of understanding the role of the human genome in disease etiology. I still maintain close contacts with Dr. Sunyaev and Mr. Francioli of the consortium and hope to continue collaboration in my graduate work.