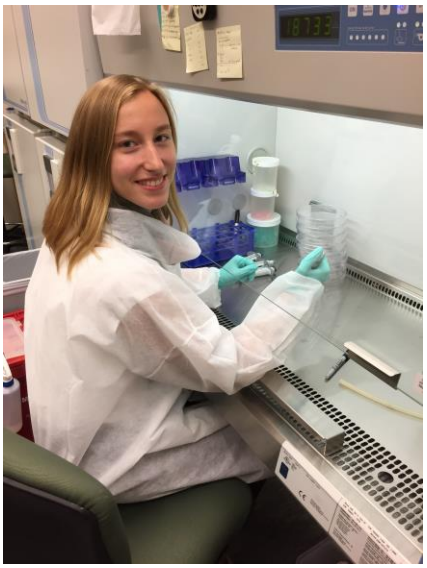

Analyzing and Visualizing High-resolution 3D Genome Folding Data Across Cell Types
Caroline Lachanski (EAS 2018)
Advisor: Jennifer Phillips-Cremins

This past summer, I had the opportunity to work in the lab of Dr. Jennifer Phillips-Cremins in the Bioengineering Department. The Cremins Lab studies the epigenetic mechanisms regulating development and function of the mammalian central nervous system; we are especially interested in the three-dimensional organization of the genome and the mechanisms that regulate it. To map the 3D organization of chromatin, the lab uses a technique called Chromosome-Conformation-Capture-Carbon-Copy (5C) along with high-throughput sequencing, which together can detect the physical interactions between different regions of the DNA. We are especially interested in finding chromatin looping interactions, concentrated points of high interaction frequency relative to the surrounding background. These "loops" often link regulatory elements of the genome, such as linking an enhancer to a gene.

We focused on three cell types: embryonic stem (ES) cells cultured in 2i media, ES cells cultured in serum, and neural progenitor cells (NPCs). For each of the three cell types, we performed 5C in six regions in the genome. For example, one region we study is the Sox2 region, which represents the Sox2 gene, a gene associated with the maintenance of pluripotency, and approximately one Mb of the surrounding DNA.

From this 5C data, we can generate contact frequency matrices for our particular regions of interest. Each ijth element of the matrix represents the number of interactions observed between the DNA fragment at the ith locus and the DNA fragment at the jth locus. These contact frequency matrices are visualized as heatmaps. However, the initial data possesses many technical biases, and the heatmaps are often difficult to interpret. Thus, our goal was to develop a computational and statistical pipeline to identify looping interactions and compare loops across cell types. Through our pipeline, we correct for technical biases, reduce spatial noise, model distance-dependence and local expected background, identify significant looping interactions, and quantitatively assess their dynamic behavior across three cellular states.

Through this research project, I gained a lot of valuable computational, biological, and analytical skills. I was exposed to Python and its uses in computational biology; I also learned more about scientific writing and presentation.

Participating in this opportunity truly exposed me to what a career in research would be like. I learned a lot about what goes into the making of a research paper, including figure creating, paper writing, and editing. This immersion into research added another dimension to my academic experience as I was able to learn concepts and skills that went well beyond the classroom.