



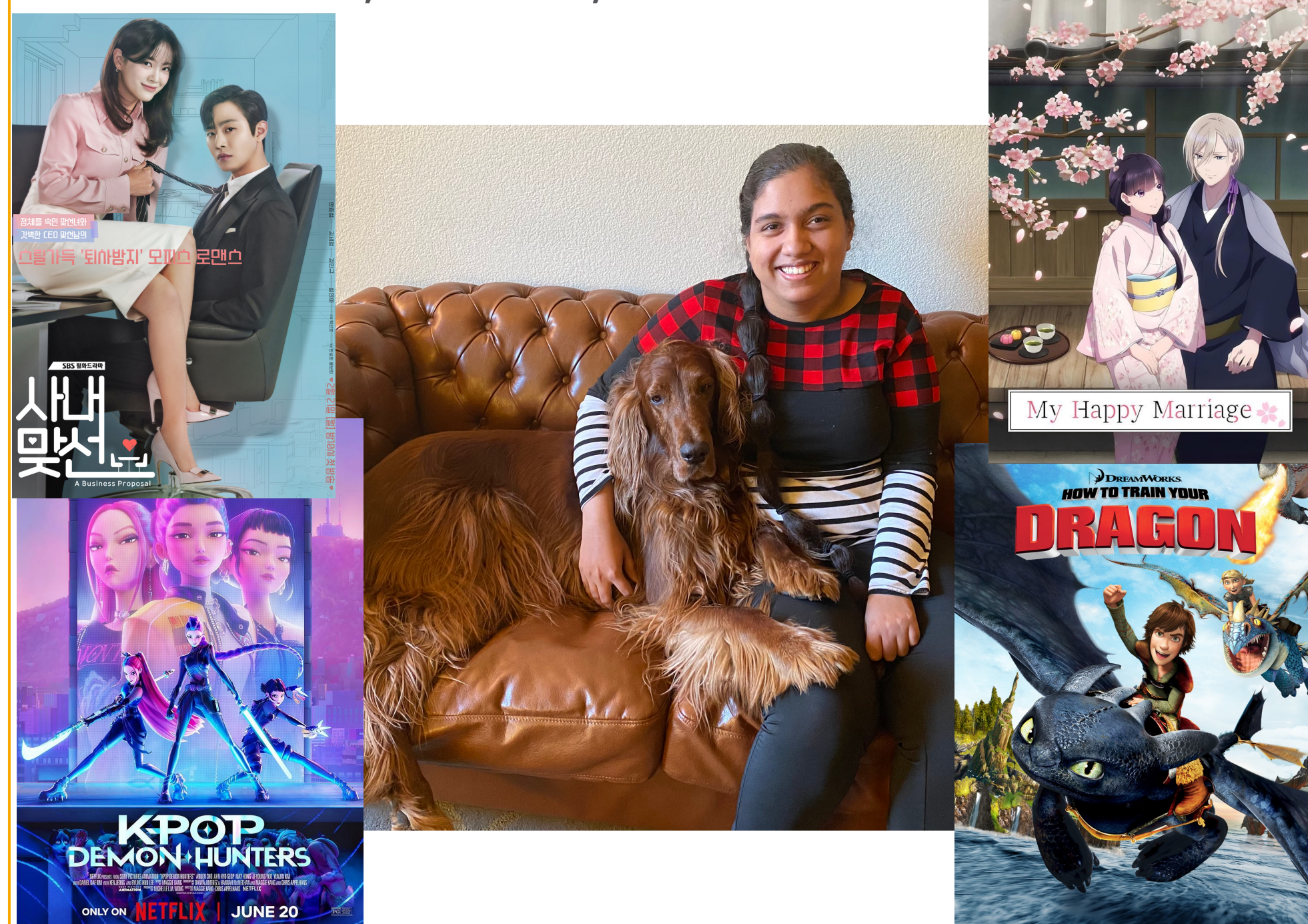
Machine Learning to Discover Genetic Structure of Amyotrophic Lateral Sclerosis (ALS)

Sanah Keswani-Santiago, University of California, Davis

Mentor: Alan Kaplan, Lawrence Livermore National Laboratory

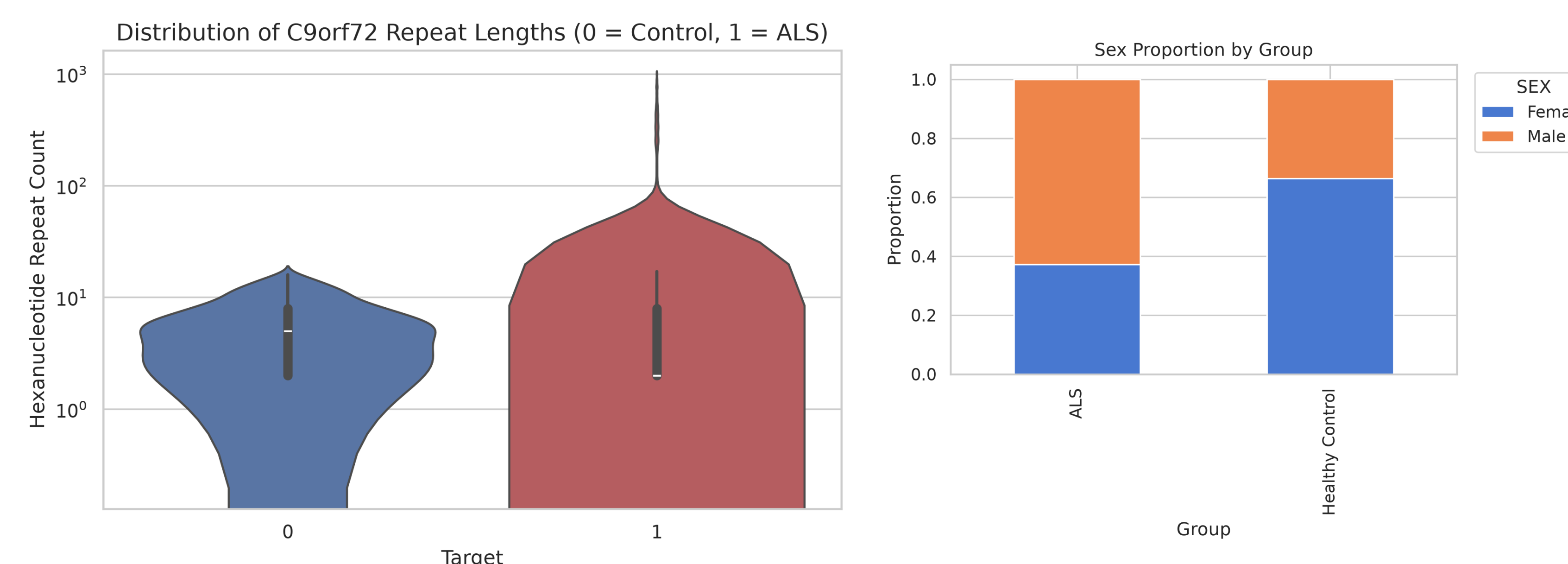
ABOUT ME

- **Majors:** Biotechnology (Bioinformatics specialization) and Data Science.
- **Career Goals:** Contribute to improving and creating treatments and cures for diseases.
- **Why LLNL?:** I chose to have a fellowship at LLNL because of the abundant resources and opportunity to perform research under guidance and mentorship at an amazing national lab.
- **Interests:** Listening to music, playing videogames, watching TV shows and movies (including K-Dramas and anime), reading books, trying different foods, collecting art and trinkets, spending time with friends, nature, animals.



SUMMER RESEARCH

- **Abstract:** Amyotrophic lateral sclerosis (ALS) is a progressive neurodegenerative disease with substantial clinical and genetic heterogeneity. Identifying genetic markers associated with ALS may improve understanding of disease biology and support development of predictive computational tools, which can help improve diagnostics and treatments, and create cures. We aim to investigate genetic markers within known ALS genes and evaluate whether demographic, clinical, and genetic features can be used to distinguish ALS cases from controls. By integrating genetic and metadata variables, we want to develop machine learning classifiers capable of separating ALS from healthy controls and to identify features that are most informative for classification.
- **Methods:** We are analyzing data from 543 ALS participants and 99 healthy controls from an AnswerALS study. Genetic, demographic, and clinical information are used to compare ALS cases with controls and to look for genetic markers, including C9orf72 repeat length. We are using scikit-learn to utilize machine learning and build a classifier that could distinguish ALS from healthy controls based on these features. Model performance is evaluated to see how well the genetic and clinical data separate the two groups.



This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.

LLF AWARD

- **My LLF Funds are being used to** help with housing, living and transportation expenses, food, groceries, and expenses related to my physical disability.
- **This fellowship has given me new and exciting opportunities to** gain applied knowledge in research and teamwork and to apply my skills to help make an impact on an important project. It has given me access to resources and enabled me to make great, meaningful connections.



LLNL-POST-2021418