Hepatocellular carcinoma (HCC), a primary form of liver cancer, is one of the deadliest cancers currently affecting humans. The most common cause of HCC is a chronic hepatitis B virus (HBV) infection. Although vaccines and antivirals for HBV exist, many individuals remain infected or vulnerable to a chronic infection and the rate of cancer development from such an infection remains high. To improve the prognosis of this cancer, we need to improve our understanding of the tumour microenvironment to develop novel, effective therapeutics. Considering the availability and experimental challenges of working with human tissue, the progression of therapeutics is often aided by a comprehensive animal model.

In my work, I present the eastern woodchuck, *Marmota monax*, as a single-cell compatible animal model that reflects key qualities of human liver disease. The woodchuck is one of the few mammalian species naturally susceptible to HCC development when infected with a close relative of HBV, the woodchuck hepatitis virus (WHV). To establish the woodchuck as a single-cell compatible model, I annotated a newly sequenced woodchuck genome which was then used to study individual cells in the woodchuck liver and blood. Cells were analyzed from both healthy and chronically infected tissue to understand the phenotype of diseased woodchuck cells when compared to a healthy reference. Diseased woodchuck liver cells exhibit many of the same characteristics of diseased human liver cells, including a significant increase in T cell exhaustion signatures.

Genome annotation, the process of identifying features on a genome sequence, was a challenging process that lacked clear guidelines in the literature. To address this issue, I created a tutorial on genome annotation that builds off existing genome annotation tools and guidelines. Additionally, when studying liver single-cell RNA-sequencing data, I found that many cells were

contaminated with high quantities of ambient RNA and lacked the expression of key nuclear genes. I determined that if a cell lacked *MALAT1* expression, the expression of a strongly expressed nuclear long non-coding RNA, above a certain threshold, then the cell likely lacked a nucleus. Such cells should be flagged or removed from a single-cell RNA-sequencing experiment.