If we delve into the world of viruses, we would find that viruses have no cells , which is quite unlike prokaryotic and eukaryotic organisms which are composed of a fully functional cell , viruses are just pieces of DNA or RNA; DNA that is de-oxy ribonucleic acid is the component that makes up the genetic material of an organism , RNA on the other hand is the component using which DNA expresses itself , ribonucleic acid or RNA is responsible for protein synthesis , now viruses are just pieces of DNA or RNA wrapped around in a protein sheet; and they cause diseases by integrating there genome sequence into the host cell and overtaking the cellular machinery ; after which they replicate and infect other healthy cells in turn causing a viral infection .

Coming to retro viruses which were the basic focus of our study ;retroviruses are a specific type of RNA viruses, which follow a process called reverse transcription to convert RNA(Ribo-Nucliec Acid) into DNA(Deoxy-Ribonucliec Acid) using the reverse transcriptase enzyme, and integrating that DNA into the host cell , these are very hard to detect and have infected more that 560 million people till date .

Hence DNA replication becomes a fundamental step in these viral diseases . DNA replication is the process by which the genetic information inside a cell is replicated and passed on to the newly formed cell which ultimately leads to the growth of the organism . DNA molecule has two strands during replication they unwind , replicate and divide . Origin of replication is the point where the DNA replication starts which leads to the multiplication and growth of viruses . If you consider the genome as a very long string of characters, these characters are technically referred to as base pairs, then the most common occurring sequence of a particular length referred to as K-Mer is the origin of replication .

The solution which we have developed is called GenHACK , and the core idea behind the application is to find the origin of replication inside the viral genome using our custom AI/ML algorithm , which can then be passed to geneticists in wet labs to slice it or inhibit it using a slicer or an inhibitor enzyme .

This is the workflow of our application , we have used the NCBI API’s to obtain the viral genome sequence from the NCBI database in FASTA format(a widely used format to exchange bioinformatics data) , from there we pass it on to our Bio-Python rest module, which performs 3 functions ,firstly finds the Origin of replication , secondly performs genomic analysis on the genome like finding the mRNA seq , amino acids and the protein end product produced by the virus , and lastly it uploads the data we gathered in the previous steps to a FHIR server , using a profiled molecular seq resource making it available for other applications to consume .

Apart from this we also have a 3D modeling functionality integrated inside our application using ICN3D API’s .

The infra architecture of our application is completely cloud based , we have used public azure to develop and deploy our application , we have used Azure Repos for version control , Azure Devops for continuous integration and continuous deployment and Azure Container Registry and Azure Kubernetes Services to store and deploy our docker images and Azure monitoring and app insights for app instrumentation .

This was the end to end flow of our application , now I would like to hand it over to Rakhee to explain the advantages and future prospects of our application .