CHAPTER 1: INTRODUCTION

1.1. Introduction

In late December of 2019, a novel respiratory disease was reported for the outbreak in Wuhan, China. It was later known after a genomic analysis as the Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) or Coronavirus Disease 2019 (COVID-19) after a genomic analysis from respiratory samples. The novelty of SARS-CoV-2 was found to have similar genetic sequence identity to the earlier known coronavirus with 79% and 50% for SARS-Cov and MERS-CoV, respectively. (Yadav et al., 2021). However, the COVID-19 began to rapidly spread worldwide due to its high transmission rates and resulted in a global pandemic (Azzi et al., 2020).

Coronavirus belongs to the family of Coronaviridae with an enveloped positive-sense single stranded RNA genome virus and a size ranging from 26 to 32 kilobases (Yadav et al., 2021). SARS-CoV-2 virus particle consists of the viral RNA containing the genetic materials as well as structural proteins. The genome includes variable amounts of Open Reading Frames (ORFs) responsible for encoding the structural proteins and non-structural proteins. The encoded proteins are essential for virus assembly, transcription, replication in order to make up the virus particle ("ORF3a protein [Severe acute respiratory syndrome coronavirus 2] - Gene - NCBI", 2021). In addition, several groups of the accessory genes are distributed, or sometimes overlapping, amongst the structural genes. The overlapping gene expressions are usually caused by nucleotide substitutions during transcription, resulting in a novel protein. The accessory proteins are studied to have additional functions required in the viral pathogenicity, for instance, the modulation of the host interferon signaling pathways (Michel, Mayer, Poch & Thompson, 2020). Although, studies are still researching on the significance of accessory proteins and how they contribute to the virus.

Due to the severity of the disease, vaccines and immunotherapeutics for SARS-CoV-2 had become an urgent demand. The SARS-CoV-2 vaccine was quick to be developed and produced in regards to the previous extensive knowledge on the vaccine development for other diseases. The Wuhan-Hu-1 genome reference sequence was used as the foundation for vaccine candidates

(Dearlove et al., 2020). The development of the vaccines included recombinant spike protein-based, whole inactivated as well as attenuated vaccines. The results showed a positive protection against the disease, however may not induce sterilizing immunity (Amanat & Krammer, 2020).

In addition, T-cell immunity plays a major role in the control of viral pathogenesis and immune response triggered by the viral proteins like spike and nucleocapsid. The viral antigens are recognized by the T-cells presented by MHC complexes (Shah, Firmal, Alam, Ganguly & Chattopadhyay, 2020). As such, the human CD4+ helper T-cells manage the immune response by enabling the production of antibodies from the B-cells, while CD8+ cytotoxic T cells help in eliminating cells infected with the virus (Nelde et al., 2020). Both T-cells recognize the viral antigens via the short peptides present on the Human Leukocyte Antigens (HLAs) (Cun et al., 2020).

The immunity mediated by T-cells contributes to the host protection against the viral infection, a rapid genome sequencing of the human T-cell epitopes in relevant to SARS-CoV-2 allows the early detection for the occurrence of variation within the T-cell epitopes (Shah, Firmal, Alam, Ganguly & Chattopadhyay, 2020). As such, modern immunoinformatics methodologies provide new strategies for the epitope-based vaccine research against SARS-CoV-2 (Cun et al., 2020). To be specific, this study analyses the identification of T-cell epitope variation in the accessory proteins between the circulating SARS-CoV-2 strains against the ancestral strain (Wuhan-Hu-1) in order to determine the relevance of the vaccines towards the SARS-CoV-2 strains circulating at the time.

1.2. Objective

To evaluate the genome analysis between the SARS-CoV-2 circulating strains and the ancestral strain using computational immunoinformatics approach. Furthermore, the thesis research project also aims to observe the variation of T-cell epitopes in circulating SARS-CoV-2

strains in comparison with the ancestral strain in order to study the relevancy of vaccines towards the SARS-CoV-2 strains circulating at the time.

1.3. Scope of Work

The scope of work for this thesis project includes the Next Generation Sequencing (NGS) on the circulating SAS-CoV-2 strains in GSI Lab, genome sequence alignment and translation to identify the accessory proteins of circulating SARS-CoV-2 strains, immuno bioinformatics approach on the analysis of the T-cell epitopes and SARS-CoV-2 T-cell epitope conservation analysis for surveillance purposes.