Arab American University
Faculty of Graduate Studies
Department of Health Sciences
Master Program in Molecular Genetics and Genetic Toxicology

Correlation Between HLA Type and Covid-19 Severity Among the Palestinian Population

Muna Mufeed Yousef Mahmoud
201911914

Supervision Committee:
Dr. Zaidoun Salah
Dr. Nouar Qutob
Dr. Kamal Dumaidi
Dr. Robin Abughazaleh

This Thesis Was Submitted in Partial Fulfilment of the Requirements for the Master Degree in Molecular genetics and genetic toxicology

Palestine, October/ 2024

Arab American University

Faculty of Graduate Studies

Department of Health Sciences

Master Program in Molecular Genetics and Genetic Toxicology.

Thesis Approval

Correlation Between HLA Type and Covid-19 Severity Among the Palestinian Population

Muna Mufeed Yousef Mahmoud
201911914

This thesis was defended successfully on 12/10/ 2024 and approved by:

Thesis Committee Members:

Name	Title	Signature
1. Dr. Zaidoun Salah	Main Supervisor	
2. Dr. Nouar Qutob	Member of Supervision Committee	
3. Dr. Kamal Dumaidi	Member of Supervision Committee	
4. Dr. Robin Abughazaleh	Member of Supervision Committee	

Palestine, October/ 2024

Declaration

I declare that, except where explicit reference is made to the contribution of others, this thesis is substantially my own work and has not been submitted for any other degree at the Arab American University or any other institution.

Student Name: Muna Mufeed Yousf Mahmoud.

Student ID: 201911914

Signature: *Muna Mahmoud*

Date of Submitting the Final Version of the Thesis: 27/5/2025

Dedication

To my beloved husband, Wael, whose unwavering support and endless encouragement have been my greatest source of strength.

To my precious family, especially my mother and my beloved children, for their love, patience, and continuous motivation.

To the soul of my late father, I pray for his eternal peace and mercy, and to the souls of all who lost their lives during the pandemic—may God grant their families strength and comfort.

Muna Mufeed Yousef Mahmoud

Acknowledgments

First, I would like to thank all who have helped me with this project; I will start by thanking Dr. Nouar Qutob and Ph. Zaidoun Salah, who were my supervisors during this project, for all their patience, support, and recommendations, as well as for all that they have taught me. Special thanks go to Dr. Fouad Zahida for his vital contributions to the data analysis, which greatly enriched the study and provided a solid foundation for my research.

I sincerely thank the Arab American University for providing the laboratory facilities, equipment, technical support, and professional staff needed to complete this project successfully.

I am also grateful to the Palestinian Ministry of Health, represented by Mr. Ossama Najjar, the Assistant Deputy Minister for Par Medical and Blood Banks, for the support and cooperation that contributed to the success of this study.

To everyone who contributed to this work in any capacity, thank you.

Correlation Between HLA Type and Covid-19 Severity Among the Palestinian Population

Muna Mufeed Yousef Mahmoud

Supervision Committee: Dr. Zaidoun Salah

Dr. Nouar Qutob

Dr. Kamal Dumaidi

Dr. Robin Abughazaleh

Abstract

Background: Coronavirus disease 2019 (COVID-19), caused by SARS-CoV-2, was declared a global pandemic by the World Health Organization (WHO) on March 11, 2020. In the West Bank and Gaza Strip, the deaths are 5,708, with over 703,228 infection cases. (University, 2020). The progression of associated symptoms varies considerably, ranging from asymptomatic cases to severe, critical, and fatal ones. (F. Wang et al., 2020a)'(WHO, 2020) This variability has prompted global research to understand the underlying factor. One major focus has been the Human Leukocyte Antigen (HLA) system and its role in shaping the immune response and influencing disease severity.

Purpose: The study aimed to explore the statistical correlation between HLA class I alleles and COVID-19 severity among a Palestinian cohort of 18 ICU patients in the West Bank with severe and critical illness during the period from February 2021 to September 2021. It also aimed to suggest a potential immunological mechanism involving T cell epitope interaction.

Methods: DNA extracted from blood samples of ICU patients was subjected to whole-exome sequencing using Next Generation Sequencing technologies. Genotyping was conducted using HISAT, with comparisons across HLA-LA, HISAT genotyping, and Seq2HLA tools. Statistical associations between HLA-I alleles and disease severity were assessed using Fisher's Exact and Yates χ^2 tests via Prism software. The Benjamini-Hochberg method was used for P-value correction. Additionally, Python scripts were developed to visualize T cell epitopes on the SARS-CoV-2 Spike and Nucleocapsid proteins using the Immune Epitope Database (IEDB).

Results: The analysis identified several HLA-I alleles A*01:03, B*15:10, B*57:03, C*03:04, C*07:01, C*07:02, and C*14:02. as significantly associated with critical COVID-19 cases ($P < 0.05$). HLA-C*07:01 and HLA-C*07:02 may play a role in disease severity through their interaction with key viral epitopes, potentially facilitating viral entry or immune.

Conclusion: These findings suggest a possible role for these HLA-I alleles in modulating the immune response and influencing the severity of COVID-19.

Keywords: Covid-19, West Bank, The Human Leukocyte Antigen System, Virus SARS CoV-2.

Table of Contents

#	Title	Page
Declaration		I
Dedication		II
Acknowledgments		III
Abstract		IV
List of Tables		VII
List of Figures		VIII
List of Appendices		IX
List of Definitions of Abbreviations		X
1.	Chapter One: Introduction	1
1.1	Background	1
1.2	Thesis Statement	3
1.3	Study Objective	3
1.4	Study Significance	4
1.5	Research Questions	4
1.6	Conceptual Scope and Variables Definition	4
1.6.1	Study Variables	5
1.6.2	Conceptual Definitions of Illness Severity	5
1.7	Defined Boundaries of the study	5
1.7.1	Temporal Boundaries	5
1.7.2	Population and Sample Boundaries	6
2.	Chapter Two: Literature Review	7
2.1	SARS CoV-2 structure and genome.	7
2.2	SARS CoV-2 Entrance to host cells	10
2.3	SARS CoV-2 replication	12
2.4	Immune response in viral infections	13
2.5	Human Leukocyte Antigen System	17
2.6	Human leukocyte Antigen alleles related to SARS CoV2	21
3.	Chapter Three: Methodology	27
3.1	Study Design	27

3.2 Study Cohort (Site and sample) 28
3.3 Ethical consideration 28
3.4 Sample collection 28
3.5 HLA Genotyping 29
3.5.1 DNA Extraction 29
3.5.2 Next Generation Sequencing 31
3.6 Statistical Analysis 34
3.7 Epitopes exploring. 35
4. Chapter Four: Results 37
4.1 Genotyping Class I HLA using three distinct software programs. 37
4.2 The frequency and odds ratio values of the HLA alleles with uncorrected significance in COVID-19 patients. 39
4.3 The significant HLA alleles among COVID-19 patients 47
4.4 The T cell epitopes for HLA alleles among Covid-19 patients. 48
5. Chapter Five: Discussion and Conclusions 51
5.1 Statistical Analysis and Results 51
5.2 Epitopes Exploring 55
5.3 Conclusions 57
5.4 Limitations 58
5.5 Recommendation 58
References 60
Appendices 80
ملخص 92

List of Tables

Table #	Title of Table	Page

Table 4.1 : The HLA alleles for COVID-19 patients were obtained by three different software. 38

Table 4.2: Hisat-genotype analysis of HLA-A alleles among COVID-19 patients displaying uncorrected p-values. 41

Table4.3 : Hisat-genotype analysis of HLA-B alleles among COVID-19 patients displaying uncorrected p-values. 41

Table 4.4 : Hisat-genotype analysis of HLA-C alleles among COVID-19 patients displaying uncorrected p-values. 42

Table 4.5 : HLA-LA analysis of HLA-A alleles among COVID-19 patients displaying uncorrected p-values. 43

Table 4.6 : HLA-LA analysis of HLA-B alleles among COVID-19 patients displaying uncorrected p-values. 43

Table 4.7 : HLA-LA analysis of HLA-C alleles among COVID-19 patients displaying uncorrected p-values. 44

Table 4.8 : Seq2HLA analysis of HLA-A alleles among COVID-19 patients displaying uncorrected p-values. 45

Table 4.9: Seq2HLA analysis of HLA-B alleles among COVID-19 patients displaying uncorrected p-values. 45

Table 4.10: Seq2HLA analysis of HLA-C alleles among COVID-19 patients displaying uncorrected p-values. 46

Table 4.11: The significant HLA alleles among COVID-19 patients by Hisat-genotype program with corrected P value. 47

Table 4.12: The significant HLA alleles among COVID-19 patients by HLA-LA program with corrected P value 47

Table 4.13: The significant HLA alleles among COVID-19 patients by Seq2HLA program with corrected P value 48

List of Figures

Figure #	Title of Figure	Page
Figure 2.1	Schematic of the SARS-CoV-2 genome and protein products.	9
Figure 2.2	Schematic diagram of HLA Genes Locus on short the arm of chromosome six: showing the three classes of HLA system.	18
Figure 3.1	Schematic of the Study Design	27
Figure 3.2 :	The Inclusion and exclusion Criteria for the study sample.	29
Figure 3.3:	The Agarose Gel Electrophoresis image of the extracted DNA…….	30
Figure 3.4 :	NGS Data Analysis Workflow.	31
Figure 3.5 :	A schematic diagram illustrating the distribution of study samples.	32
Figure 4.1 :	The T cell epitopes for significant HLA alleles in Table 4.11.	49
Figure 4.2:	The T cell epitopes for significant HLA alleles in Table 4.11.	50

List of Appendices

Appendix #	Title of Appendix	Page
Appendix A:	The Study Sample Distribution and patient's Clinical test results.	80
Appendix B:	DNA Concentrations.	82
Appendix C:	Spike protein HLA Epitopes.	83
Appendix D:	SARS COV2 Nucleocapsid protein epitopes.	89
Appendix E:	Python Code Snippet	90
Appendix F:	IRB Approval	91

List of Definitions of Abbreviations

Abbreviations	Title
3CLpro	3C- like protease
Ang-II	Angiotensin II
ACE	Angiotensin-converting enzyme
ACE-2	Angiotensin-converting enzyme-2
Ang-I	Angiotensin I
APC	Antigen-presenting cells
ARDS	Acute Respiratory Distress Syndrome
Bcl-xL	B-cell lymphoma-extra large
C3a	Complement 3 a
CBC	Complete blood count
CDC	Centers for Disease Control
CH	Central helix
CK	Creatine Kinase
CNX	chaperone calnexin
COV	Coronavirus-1
COV-2	Coronavirus-2
COVID-19	Coronavirus disease 2019
CRP	C Reactive Protein
CSSE	Center for Systems Science and Engineering
CTD	C-terminal domain
E	Envelop
eBLT	Enrichment-based linked transposons
ER	Endoplasmic reticulum
ERAP1	Endoplasmic reticulum aminopeptidase 1
ERAP2	Endoplasmic reticulum aminopeptidase 2
FCoV	Feline Coronavirus
FiO2	fraction of inspired oxygen
FP	Fusion Peptide
HLA	Human Leukocytes Antigen
HR	Heptad repeats
ICTV	International Committee on Taxonomy of Viruses
ICU	Intensive Care Unit
IFNs	Interferons
IL	Interleukins
ISGs	Interferons stimulated genes
JHU	Johns Hopkins University
LDH	Lactic dehydrogenase

LKR	linker region
LPV	Lung-protective ventilation
M	Membrane
MCH	Major histocompatibility complex
MDSC-like	Myeloid-derived suppressor-like
MERS	Middle East Respiratory syndrome
MHC	Major histocompatibility complex
MOH	Ministry of Health
N	Nucleocapsid
NIV	Non-invasive ventilation
NSP	Non-Structural protein
NTD	N-terminal domain
PALS1	Protein Associated with LIN7 1
PaO2	Arterial partial pressure of oxygen
PLpro	Papain- like protease
pp1a	Polyprotein 1a
pp1b	Polyprotein 1b
pMHC	MHC-peptide complex
RAS	Renin–angiotensin system
RBD	Receptor binding domain
RdRp	RNA dependent polymerase
RNAi	RNA interference
RNP	Ribonucleoprotein
RR	Respiratory rate
RTCs	Replication Transcription Complexes
S	Spike
S1	subunit 1 in spike protein
SARI	Severe Acute Respiratory infection
SARS-CoV	Severe acute respiratory syndrome Coronavirus
SARS-CoV-2	Severe acute respiratory syndrome Coronavirus two
SpO2	Oxygen saturation
TAP	Transporter associated with antigen processing
TCR	T Cell Receptor
Th	T helper cell
TLR	Toll-like receptor
TMD	Transmembrane domain
TNF	Tumor Necrosis factor
WHO	World Health Organization

Chapter One: Introduction

1.1 Background

On December 31, 2019, the WHO was informed officially of unidentified pneumonia cases of unknown etiology in Wuhan City, China (World Health Organization, n.d.). later, on March 11, 2020, it was declared it is the COVID-19 pandemic, more than 676,609,955 infections and at least 6.88 million deaths have been reported worldwide (University, 2020). This has made the COVID-19 pandemic the most significant societal concern globally according to the WHO and CDC.

Palestine was among the earliest countries to raise an emergency state; President Mahmoud Abbas declared it on 5th March 2020 as the first case of COVID-19 infection was detected in Bethlehem (State of Palestine, 2020). The West Bank and Gaza Strip have experienced multiple waves of COVID-19 infections with 5708 death cases and over 703228 infected cases (University, 2020).

Different variants of the Sever Acute Respiratory syndrome Coronavirus 2 (SARS-CoV-2) were transmitted to the country; An early Palestinian genomic study showed that the most common SARS-CoV-2 genomes are related to the single cluster B.1.1.50 during the period from Feb 2020 to Aug 2020, and other clusters to a less ratio, indicating that the Palestinian genomic diversity of SARS- CoV-2 reflects the global diversity of SARS- CoV-2 (Qutob et al., 2021). In Sep 2020, a new rapid transmission variant emerged in England, it was transmitted to the Palestinian-occupied territories in early of Feb 2021; the Palestinian MOH declared that sequencing of 139 samples revealed the presence of the B.1.1.7 (Alpha) variant.

Later on, in June 2021, the Delta variant was documented in India, and it was transmitted to the Palestinian-occupied territories. At the end of 2021, the Palestinian MOH announced 53 infected cases with the new variant Omicrons, the wide world spread variant as it reached an 11% spread rate (World Health Organization, n.d.) (Palestinian MOH, n.d.).

Palestinian COVID-19 patients reflecting the same global situation, show a consistent range of symptoms and disease manifestations, including hypoxia, respiratory distress, thrombosis, heart failure, kidney failure, and multi-organ dysfunction due to mal- immune response guided by a devasting cytokine storm.

The Palestinians' COVID-19 health management protocol (Palestinian National Scientific Committe, n.d.) has classified COVID-19 patients into three classes based on the severity of their illness. The majority, accounting for approximately 81% of cases, fall under the category of uncomplicated illness, comprising both mild and moderate forms. Severe illnesses, constituting about 14% of cases. Critical illness, comprising approximately 5% of cases (Palestinian National Scientific Committe, n.d.).

While there is no specific cure for COVID-19, treatment in Palestinian hospitals focuses on symptomatic therapy and vital sign monitoring including chest imaging and a set of routine laboratory tests to evaluate the disease manifestation (CBC, CRP, liver enzyme, myocardial enzyme, renal function), measuring arterial blood gas, checking Ferritin, fibrinogen, LDH, and others as necessary, Notably, neutrophils counts elevation that indicates a secondary infection considered a remarkable sign of severe illness consequences. Also, elevated D-dimer levels exceeding 5000 units pointed to a potential pulmonary embolism, while monitoring creatinine levels serves as an indicator of renal dysfunction. Additionally, elevated levels of creatine kinase (CK) and lactate dehydrogenase (LDH) may signify underlying heart failure.

Supportive care manages the symptoms and sustains organ function in severe cases, this includes immediate oxygen therapy to maintain oxygen saturation above 94% for hypoxic patients, severe acute respiratory illness (SARI), and those experiencing respiratory distress. The O2 supportive therapy starts at 5 L/min and adjusts the flow rates until SpO2 $\geq$ 93%, while the target is > 90% via face mask with reservoir bag (at 10–15 L/min) in critical illness patients. If the Patients show signs of severe respiratory distress or hypoxemia, a high-flow nasal cannula with 50–60 L/min flow rates may be used. If the hypoxemia persists and SpO2 < 90%, even with the high flow O2, mechanical ventilation should be instituted early. Consider Non-invasive ventilation (NIV) in case of ARDS without decreased consciousness or cardiovascular failure; if NIV is unsuccessful, do not postpone endotracheal intubation, NIV is a bi-level positive airway pressure through a tight-fitting mask, while Invasive mechanical ventilation directed through an end tracheal tube or tracheotomy.

The target in ARDS patients is to maintain a low-volume, low-pressure ventilation (LPV) strategy via Invasive mechanical ventilation with a tidal volume of 6 ml/kg body weight (Palestinian National Scientific Committe, n.d.).

Appendix A includes the sample study patients' lab test results and symptoms in the appendix chapter.

The clinical heterogeneity observed among COVID-19 patients—ranging from mild symptoms to critical illness or death—has raised the possibility of underlying genetic factors influencing the immune response. Emerging global studies have supported this hypothesis, suggesting that host genetic variation may partially explain the differences in disease severity. Nevertheless, the precise mechanisms remain unclear, highlighting a gap that this study seeks to address.

1.2 Thesis Statement

The HLA system is known to have a strong correlation with the severity of Coronavirus diseases such as Middle East Respiratory syndrome (MERS) and Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV), also, it is known to be associated with certain autoimmune diseases. SARS-CoV-2, which can be fatal in some cases due to an imbalanced immune response, HLA system has been the focus of recent global studies that suggest COVID-19 severity is potentially influenced by HLA-I alleles. These alleles vary significantly across different ethnic and geographic populations. Despite the extensive global research on the genetic factors influencing COVID-19 severity, there has been no focused study in Palestine. Given the genetic diversity and distinct HLA profiles observed in different populations, this study aims to investigate the relationship between HLA-I alleles and COVID-19 severity among hospitalized COVID-19 patients in Palestine using advanced genetic techniques such as NGS. This study will contribute valuable population-specific genetic data and lay the groundwork for future genetic research in viral infection in Palestine.

1.3 Study Objective

This study focuses on the correlation between HLA type and COVID-19 severity among Palestinian ICU and hospitalized patients with no documented history of any pre-existing diseases, chronic conditions, or significant health-related abnormalities. Aiming to genotype HLA-I profiles via WES by applying advanced NGS technique.

The study aims to: 1. Identify the correlation between HLA-I types and the severity of COVID-19 among the Palestinian population. 2. Identify specific HLA-I alleles significantly

associated with COVID-19 severity in the context of the ethnic and geographical characteristics of Palestinians, as compared to other populations, due to their unique genetic makeup. Indicate potential HLA epitopes for future COVID-19 vaccine studies.

1.4 Study Significance

This study provides genetic insights and practical applications of advanced molecular techniques that can improve health outcomes for Palestinian patients and contribute to the scientific understanding of COVID-19 in the context of the Palestinian population. The significance of this study is its contribution to the scientific understanding of the role of HLA-I alleles in infectious diseases. Providing Population-Specific Genetic Data can be used to compare with other populations globally. Improving vaccine efficacy by identifying specific HLA-I alleles associated with severe COVID-19. Enhance the development of public health strategies; such as priority of vaccination to certain categories. As the first study of its kind in Palestine, it lays the groundwork for future research on genetic factors associated with viral disease severity in the Palestinian population.

1.5 Research Questions

1. Is there a statistically significant correlation between HLA class I alleles and the severity of COVID-19 in Palestinian hospitalized patients?
2. Which specific HLA-I alleles are most strongly associated with severe or critical COVID-19 cases among Palestinian patients?
3. How do the distributions of HLA-I alleles in the Palestinian population compare with global data in relation to COVID-19 severity?
4. Can certain HLA-I alleles be considered predictive markers for COVID-19 severity in patients with no prior chronic conditions

1.6 Conceptual Scope and Variables Definition

The research examines the potential association between Human Leukocyte Antigen (HLA) Class I genotypes and the severity of COVID-19 in critically ill Palestinian patients. Thematically, the study focuses on genetic predisposition, excluding consideration of

environmental, behavioural, therapeutic, or socioeconomic factors. It does not address viral variants, treatment protocols, focusing solely on immunogenetic markers as possible contributors to disease progression.

1.6.1 Study Variables

- **Dependent Variable**: The severity of COVID-19 illness, categorized as severe or critical, based on clinical presentation, ICU admission, need for mechanical ventilation, or patient death.
- **Independent Variable**: The presence of specific HLA Class I alleles, hypothesized to influence the immune response to SARS-CoV-2 infection.
- **Control Variables**: Demographic and clinical variables such as age, gender, presence of underlying conditions (e.g., diabetes, hypertension), and familial relationships, which may influence genetic patterns and disease progression.

1.6.2 Conceptual Definitions of Illness Severity

- Mild: No radiological evidence of pneumonia, Moderate: Fever, respiratory symptoms, and mild pneumonia, Severe: Requiring oxygen support; defined by respiratory rate ≥30 breaths/min, oxygen saturation ≤93% on room air, PaO_2/FiO_2 ≤300 mmHg, and radiological progression of lung lesions exceeding 50% within 24–48 hours, and Critical: Requiring ICU care and mechanical ventilation due to respiratory or multi-organ failure (Palestinian National Scientific Committee, n.d.).

1.7 Defined Boundaries of the study

1.7.1 Temporal Boundaries

Data and blood samples were collected retrospectively between February 2021 and September 2021, during a peak period of COVID-19 transmission in the Palestinian territories. This timeframe coincided with the dominance of the Alpha and possibly Delta variants. However, variant genotyping was beyond the scope of this research.

1.7.2 Population and Sample Boundaries

The study sample comprised 18 critically ill Palestinian patients, all under the age of 55, with no documented chronic diseases or significant pre-existing conditions. The cohort included 15 ICU patients, of whom 7 died, and 3 patients receiving high-flow oxygen therapy (≥40 L/min).

This strict selection was designed to minimize clinical heterogeneity and thereby enhance the validity of analysing potential associations between HLA-I alleles and disease severity.

Chapter Two: Literature Review

2.1 SARS CoV-2 structure and genome.

The novel coronavirus represents the seventh strain known to infect humans according to the International Committee on Taxonomy of Viruses (ICTV). Coronaviruses have infected humans in a zoonotic manner since 1965 (Malik, 2020), but this novel CoV-2 shows human-to-human transmission (W. Wang et al., 2020), increasing its incidence by an exponentially 2,6 rate. It is a spherical shape with spikes like a crown virus, it belongs to the Nidovirales order. Coronaviruses are classified into four genera: Alpha, which includes HCoV-229E and HCoV-NL63; Beta, which includes HCoV-OC43, SARS-CoV, MERS-CoV, SARS-CoV-2, and HCoV-HKU1; Gamma; and Delta (Ye et al., 2020) (Malik, 2020) (Naserghandi et al., 2020).

SARS-CoV-2 was sequenced and submitted to the GeneBank databases by 17th January 2020 with accession no (NC_045512.2); it has 29903 bp and is called severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1, complete genome. The first version has the exact accession number but was formerly called 'The Wuhan seafood market pneumonia virus (2019-nCoV). Later on, on 11th February 2020, it was called SARS-CoV2 by the (ICTV); because the genome sequence is similar to SARS-CoV with a rate reached 86.9% (Ye et al., 2020)'(Malik, 2020) (Naserghandi et al., 2020).

The SARS-CoV-2 genome is about 29.8-29.9 kb surrounded by a nucleocapsid (Khailany et al., 2020). This genome codes four structural proteins, in addition to six accessory proteins of the 3'end are known to consist of 1\3 of the genome. The structural proteins are envelope protein (E), spike protein (S), nucleocapsid protein (N), and membrane protein (M) (H. Yang & Rao, 2021). The remaining two-thirds of the 5' end genome has two open reading frames, ORF1a and ORF1b: by −1 frameshift, change the translation and encode two intersecting polyproteins pp1a and pp1ab. These polyproteins, by multiple events of cleavage mainly driven by a 3C- like protease (3CLpro) the NSP5, and to a lesser extent, by papain-like protease (PLpro), the NSP3, are processed to produce viral RNA-dependent polymerase (RdRp) and another fifteen types of polyproteins used in viral replication and transcription. Other accessory proteins are encoded by accessory genes such as ORF3a, ORF6, ORF7a, ORF7b, ORF8, and ORF10 (Ye et al., 2020) (Naserghandi et al., 2020) (H. Yang & Rao,

2021) (Khailany et al., 2020). Some of these proteins' functions are well known, while others are still ambiguous. SARS-CoV-2 is a positive-sense RNA virus; the host cell ribosomes recognize its RNA as a messenger RNA, its 5' has a cap, and its 3' end has multiple poly-(A).(Ye et al., 2020) (H. Yang & Rao, 2021).

The S gene coded the S protein which is a homo-trimeric heavily glycosylated protein. It contains about 1,200 residues responsible for the viral characteristic surface crown appearance. It is also responsible for the attachment to the host cell membrane receptors. S protein has two conformational states (prefusion and postfusion); it can be cleaved by a protease into S1 and S2 subunits; however, they persist related with each other over non-covalent bonds, this facilitates the fusion between the host cells and viral membrane. It has two main domains: a receptor-binding domain (RBD) on the upside and a fusion domain on the lower one. In contrast, it has a trans-membrane hydrophobic peptide in the middle that invades the host cell membrane when the entrance occurs by forming an embedded pre-hairpin-like structure (H. Yang & Rao, 2021) (Gupta et al., 2020).

The RBD in the S1 has two distinct conformational states: the closed 'down' state and the less stable open 'up' state; S1 has an N-terminal domain (NTD), while the S2 subunit contains four conserved structural domains. One of them is the fusion peptide. It also has two heptad repeats (HR1, HR2) and a transmembrane domain. Also, it has a central helix (CH), connector domain, and cytoplasmic tail (H. Yang & Rao, 2021) (M. Y. Wang et al., 2020) (Hardenbrook & Zhang, 2022).

The E gene is the second gene that codes a structural protein in the viral genome; it is about 228 nucleotides in length and is translated to the viral envelope that forms the outer capsid. It is a small protein with almost 76 to 109 amino acids. Nearly 25 amino acids constitute a transmembrane domain (TMD) in the hydrophobic domain. E protein structure in SARS-CoV2 still needs to be solved, but in an estimation with SARS-CoV E protein, it is believed that (TMD) forms a transport ion channel, also E protein has a charged cytoplasmic tail. The hydrophobic domain is rich in valine and leucine residues.

The E protein is processed by different post-translational modifications that guide the protein subcellular trafficking and interaction. Specific points in the post-translational modifications may form a therapeutic target as they affect the function of the E protein and the viral assembly; this is supported by the fact that not quite a lot of mutation occurred in

the E gene among the different CoV-2 variants. Furthermore, E protein participates in triggering the host inflammasome; some host cell proteins are known to interact with it, such as Bcl-xL, PALS1, syntenin, (Na+/K+) ATPase α-1 subunit, and stomatin (H. Yang & Rao, 2021) (Hardenbrook & Zhang, 2022) (Gorkhali et al., 2021).

The M gene coded the third structural protein is about 669 nucleotides long, coding 222 amino acids. The membrane protein structure still needs to be solved for both SARS-CoV1 and SARS-CoV2; it is known to contain three transmembrane domains lined by a small glycosylated amino-terminal domain. The cytoplasmic domain has a carboxy-terminal tail, Which may help in Protein localization and interaction for the viral assembly.(Gorkhali et al., 2021).

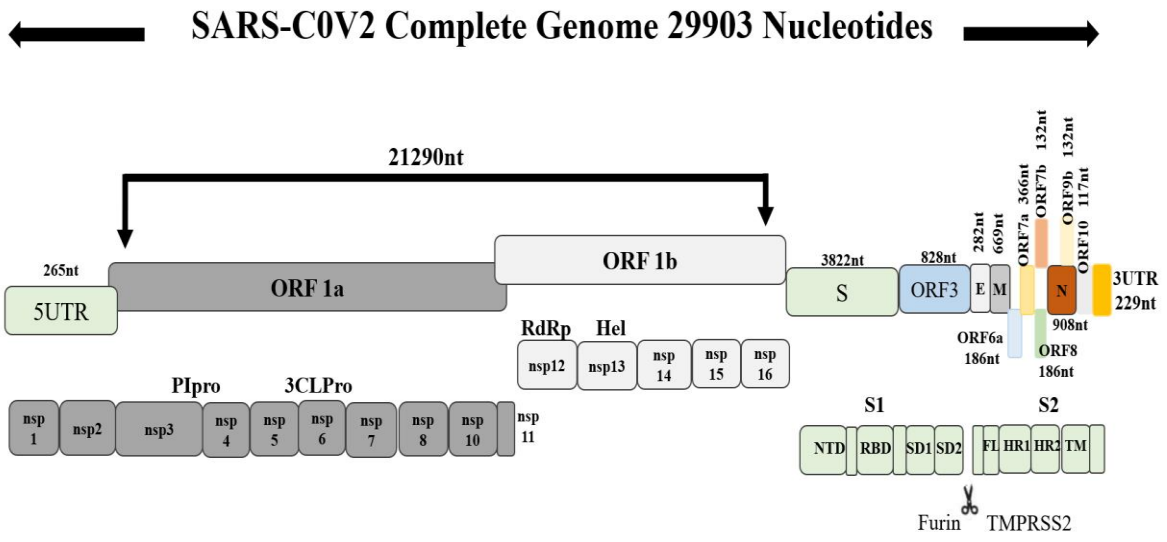

Figure 2.1 Schematic of the SARS-CoV-2 genome and protein products.
ORF1ab gene formed almost 2/3 of the 5' end virus genome; the viral genome encodes four 3' end structural proteins, and the other non-structural proteins genes formed 1/3 of the viral genome.

The N gene is about 908 nucleotides long, coding the fourth structural protein (viral nucleocapsid); it is the major viral protein in the virion; it protects the viral ribonucleotides and keeps the RNA stable inside. Also, it is essential for binding the genomic RNA. It connects to neighboring by packaging RNA into the ribonucleoprotein (RNP) complex in a

(beads on a string) mechanism. N protein can also suppress RNA interference (RNAi). It contains three disorderly domains: the N-arm, C-tail, and a middle linker region (LKR). In SARS-CoV and SARS-CoV2, the NTD works as the RNA-binding domain, and the CTD is responsible for oligomerization. SARS-CoV-2-N-NTD crystal structure was solved; it consists of four monomers, each resembling the right hand (finger, palm, and wrist) holding a pocket inside it in the packaging mode (H. Yang & Rao, 2021) (Hardenbrook & Zhang, 2022) (Gorkhali et al., 2021)

2.2 SARS CoV-2 Entrance to host cells

Angiotensin-converting enzyme-2 receptors (ACE-2) are known to be the special entrance for SARS-CoV. It is homologous to the human angiotensin 1 converting enzyme (ACE); they both belong to the angiotensin-converting enzyme family of dipeptidyl carboxypeptidases. ACE-2 is a crucial renin–angiotensin system (RAS) enzyme. It is known to be expressed in a cell-specific expression manner; they are highly expressed in lung, kidney, and heart cells, indicating that it has a role in the regulation of cardiovascular and kidney function (electrolyte and fluid homeostasis), and blood pressure. It is believed that ACE-2 is responsible for the development of cardiovascular diseases like hypertension, myocardial infarction, and heart failure. (Silhol et al., 2020)'(Renhong Yan, Yuanyuan Zhang1, Yaning Li, Lu Xia, Yingying Guo, 2020)'(Kai & Kai, 2020)

Angiotensinogen substrate cleaved by renin to form angiotensin I (Ang-I), (Ang-I) turn into angiotensin II (Ang-II) by removing two amino acids from the carboxyl-terminal by ACE. Ang-II has three types of receptors with similar affinities. When angiotensin type 1 receptor (AT1R) binds to (Ang-II) vasoconstriction, inflammation, blood coagulation, and extracellular matrix modification may occur, while angiotensin type 2 receptor (AT2R) counteracts the effect of AT1R, the third type receptor for Ang-II is ACE2 receptors (Silhol et al., 2020) (Kai & Kai, 2020).

Because of the similarity of SARS CoV and SARS CoV-II genomes, ACE2 was identified as SARS-CoV-2 human cell entrance also; the RBD in the S1 subunit of spike protein is recognized by the extracellular peptidase domain of the ACE2 receptors primarily via polar residues.(Renhong Yan et al., 2020) Also, a study mentioned that key residue substitutions in the carboxyl terminal domain of SARS CoV-2 Spike protein strengthen the

interaction and increase receptor affinity for binding.(Wang Q, et al . , 2020) Another study indicates some mutation in the S1 subunit as N501Y that pushed into the cavity near the binding domain and interacted with Y41 of ACE2, making more closeness for the two residues, and that explains the increasing affinity of binding for the Alpha variant, which has this mutation and illustrate the increase transmissibility of this variant.(Zhu et al., 2021)

Structural analysis study identified (Thr333 – Gly526) residues of the SARS-CoV-2 RBD and (Ser19 – Asp615) of the ACE2 as interaction binding sites.(Lan et al., 2020) Interestingly, RBD is believed to be the most mutational part of CoV2. The RBD has two conformational states: 1) the closed 'down' state, where the RBD borders bulled to the trimer central cavity and cover the receptor-binding site. 2) The less stable open 'up' state: where the RBD undergoes a conformational change, revealing its receptor-binding site causes regions to recognize the human ACE2 receptor on the host cellular membrane. (Del Valle et al., 2020)

Once the RBD domain of the S protein binds to the ACE2 receptor, the S2 subunit is subjected to conformational change to bring the fusion peptide and transmembrane domain together. Hence, inserting the fusion peptide into the host membrane is possible. The two heptad repeats HR1, HR2, and CD act together, forming a stranded coiled-coil bundle that results in fusion and definitive insertion of the viral genome into the host cell cytoplasm, and it is believed that the glycans protect the S2 subunit from antibody recognition by masking it at that time.(Hardenbrook & Zhang, 2022) (Cai Y, Zhang J, Xiao T, Peng H, Sterling SM, Walsh RM & Rits-Volloch S, 2020).

Then two cleavage events are essential to complete the fusion insertion process; first, RDB cleavage from the fusion domain, then a proteolytic enzymatic cleavage occurs by a cellular trans-membrane protease (TMPRSS2) to expose the fusion peptide in a viral spikes activation process.(Malik, 2020) Although some studies documented that activation can be mediated by additional proteases such as TMPRSS11A, TMPRSS11D, and TMPRSS11E1 in vitro, TMPRSS2 is still the key protease for cell entry pathogenesis (Hoffmann et al., 2020) (Mollica et al., 2020) (Hoffmann et al., 2021). Also, Hoffmann and his teams approved that TMPRSS2 inhibitor can block the viral entry and may be a therapeutic target in COVID-19 illness (Hoffmann et al., 2020). They tested the Camostat Mesylate as a protease inhibitor to block the host cell protease TMPRSS2, a drug used for pancreatitis treatment in Japan. They

found that a dose (600 mg/day) effectively inhibits COVID-19 infection of lung cells In vitro (Hoffmann et al., 2021).

2.3 SARS CoV-2 replication

SARS-CoV2 uses several host cell factors to promote a cytoplasmic replication cycle, incorporating various techniques to balance viral gene expression on its translation and transcription levels (Ziv et al., 2020). Once the virus succeeds in releasing it is genomic RNA inside the host cell cytoplasm, the single-stranded vRNA serves as mRNA and recruits the host cell ribosomes. It has at least six ORFs, and the translation process of ORF1a yields a large poly protein (PP1a); at the C terminal of this ORF, a translation continues with new ORF1b by a programmed -1frameshifting to yield a larger poly protein (PP1ab), the frameshift efficiency is about 57% ± 12%. (Ziv et al., 2020)'(Finkel et al., 2021)Furthermore, these polyproteins cleavage into 16 nonstructural proteins (NSPs) by 2 proteases, the main (Mpro) of NSP5, also identified as 3C-like protease (3CLpro), and the (PLpro) of NSP3. The pp1a produces 11 NSPs (from 1 to 11), while the pp1ab produces all NSPs except NSP11, where the frameshifting occurred and changed the ORF (Malone et al., 2022).

As soon as NSP1 is produced, it interacts with the 40S ribosomal subunit to stop host mRNAs translation via endonucleolytic cleavage close to the 5'UTR that triggers mRNA degradation, and so it blocks the innate immune response of the infected host cells. On the other hand, viral mRNA has a 5'-end leader sequence protecting it from degradation (Gorkhali et al., 2021).

The he first product yielded by automatic cleavage is Nsp5 (3CLpro); it then cleaves the polyproteins at 11 sites, producing NSPs (4 -16). NSP12, NSP7, and NSP8 work together to perform efficient RNA synthesis. The N-terminal of nsp14 has the role of proofreading 3′-5′ exonuclease to prevent fatal mutagenesis. Other NSPs, NSP3, NSPs4, and NSP6 of SARS-CoV, are needed to develop double membranes (Gorkhali et al., 2021). The other NSPs NSP12, NSP13, NSP14, and NSP16 contain enzymes involved in the replication–transcription complexes (RTCs) for viral RNA synthesis. RTCs have complexes of NSPs that have not yet been identified precisely (Gorkhali et al., 2021) (Malone et al., 2022).

When a cell is infected with SARS-CoV2, the virus prompts a vast alteration of endoplasmic reticulum (ER) membranes. One of these alterations is the production of

unusual double-membrane vesicles (DMVs); the virus takes advantage of these vesicles and assembles the RTCs in the DMVS to produce new gRNAs and start encoding the four structural proteins in addition to some accessory proteins by transcript sub-genomic RNA containing the (ORFs) 2–9b. The new gRNAs, in turn, repeat the virus translation transcription cycle, producing more NSPs and synthesis RNA (Malone et al., 2022) (Hartenian et al., 2020).

The virion assembly begins when the nucleocapsid proteins coat the new gRNA, and then the other structures obtained by budding into the endoplasmic reticulum–Golgi intermediate compartment (ERGIC) that already contains the structural proteins spike, membrane, and envelope. Lastly, these virions are released out of the infected cell by exocytosis (Hartenian et al., 2020) (Abu et al., 2020).

2.4 Immune response in viral infections

Although human innate immunity is not specialized, it is considered the first barrier against viral infection. As any virus or its particles have been recognized by the immune system, the infected cells release the humoral components, interferons (IFNs), and pro-inflammatory cytokines into the neighboring environment. These components, in turn, stimulate the maturation of naïve T cells into different subclasses Th1, Th17, Th2, and Treg (Kudlay et al., 2022) (Howard et al., 2022).

The Th1cell regulates the INF-*Y* and IL-2, while Th17 cells regulate the IL-17A, IL-17F, and IL-22. Together, they guide the pro-inflammatory response of the innate immunity via its cellular part, Which promotes more antigen-presenting cells in a positive feedback loop to the area, this aims to eliminate the virus by macrophages and to localize the viral particles by releasing inflammatory cytokines, tumor necrosis factor-α (TNF-α) and INF. (Kudlay et al., 2022) (Augusto & Hollenbach, 2022). These INFs, in turn, activate the Natural killer that kills the infected cells, leading to rapid eradication of the infection and firing the adaptive immunity via the activation of the cytotoxic T cells CD8+ cells and maturation of the B lymphocyte into a plasma-specific antibody releasing cells (Kudlay et al., 2022) (Augusto & Hollenbach, 2022).

The process of B cells maturation into plasma cells is directed by the Th2 cells; Th2 cells regulate IL-4, IL-5, and IL-12. This process also needs the T helper cells CD4+ T cells

(Howard et al., 2022). B cells also maturate to B memory cells that reactivate the plasma cells in case of reinfection, providing a rapid response and long immunity for a specific antigen (Kudlay et al., 2022). In contrast, T regulatory cells maintain immune homeostasis via IL-10 balance (Howard et al., 2022).

It's important to mention that the immune response susceptibly to viral infection differs between individuals, depending on diverse factors. Some of these factors are genetic, such as their human leukocyte antigens HLA Haplotypes, IFN-Y3 polymorphisms, miRNA in the host cell as the virus uses it in its replication, and killer-cell immunoglobulin-like receptors. Nongenetic factors may include age, having a specific disease, taking some medication, weight, and nutrient status (Howard et al., 2022). One of the remarkable viruses that have a considerable variation in the immune response is SARS-COV2. While it may be a fatal disease in some individuals, it can be asymptomatic in others. Scientists believe that a complete understanding of the COVID-19 immune response might be achieved by taking advantage of the genome similarity between SARS-COV, MERS, and SARS-COV2.

SARS-COV, also a member of the Coronaviridae family, caused a respiratory illness outbreak in 2002 – 2004 (Malik, 2020) (Naserghandi et al., 2020), showed a short-life circulating antibody and more robust T-cell immunity.

SARS-COV T cell epitopes were documented in recovered patients. The overlapping epitopes capable of causing an immune response were studied along the virus genome. Also, in recovered patients, the memory T cells that have activity against the two structural proteins of the virus S-protein and N-protein were detectable for years, even up to eleven years in some patients.(Kudlay et al., 2022)'(Fan et al., 2009)'(L. T. Yang et al., 2006)'(Ng et al., 2016) Subsequently, it was hypothesized that SARS-CoV-2 may exhibit a similar immune response.

The MERS, another member of the Coronaviridae family, caused the Middle East respiratory syndrome outbreak in 2012.(Malik, 2020) Initially, the short-life neutralizing antibody titers were directly correlated to the level of protection of the fatal illness. So, the low antibodies titer or their delayed production was linked to severe illness and death. However, In 2016, Corman showed that antibodies do not cause virus elimination.(Kudlay et al., 2022)'(Corman et al., 2015)'(Zhao, 2017)T cells (CD4+, CD8+) were detectable even in the absence of specific antibodies in MERS.(Zhao, 2017) While a significant reduction in

the T lymphocytes, IL-12 and IFN-γ levels was documented in the patient with fatal illness.(Min et al., 2016)'(Faure et al., 2014)

Regarding SARS-COV2, the antigen-presenting cells play an important role in innate immunity; macrophages and the dendritic cells underneath alveolar epithelium move towards the lymph nodes, so the viral epitopes are presented to CD4+ T and CD8+ T cells. Macrophages express a high level of ACE2; ACE2 is considered the entrance receptor for SARS-COV2 to the host cell, while the dendritic cells show no ACE2 expression. Dendritic cells identify the virus through their Toll-like receptor (TLR), which can sense the GU-rich RNA sequences and activate INFs production. INFs may cause uncontrolled inflammation in severe cases of patients. Constant with these data, a TLR7 genetic deficiency mouse model was linked to the severe illness (Salvi et al., 2021).

Another receptor on the dendritic cells proposed to interact with the N glycosylated motif of the spike protein positioned out the RBD and has a crucial role in viral entry is the C-type lectins like receptor CLR. The interaction of these receptors is associated with a hyperinflammatory response (Marongiu et al., 2021) (Xing et al., 2020), CD4+ T cells are activated by the major histocompatibility complex (MHC) class II found on the surface of these antigen-presenting cells. Then, it activates B cells to produce antibodies (Kudlay et al., 2022) (Augusto & Hollenbach, 2022). In contrast, CD8+ T cells are activated by the major histocompatibility complex (MHC) class I, founded on all nucleated cells' surfaces. Then, they activate the cytotoxic cells that can kill viral-infected cells (Kudlay et al., 2022) (Augusto & Hollenbach, 2022).

Recent studies of SARS-CoV-2 show similar antibodies production to the other coronaviruses. IgM antibodies production started within two weeks of infection, and their levels decreased about three months after infection, while IgG levels were more stable, they began to rise within two months of infection and remained for five months, according to the Wajnberg et al. study (Wajnberg et al., 2020).

A Prospective Study of SARS-CoV-2–specific antibodies in Muscovite patients showed the defensive role of antibodies and cellular responses correlating with an increase in their titers over time (Molodtsov et al., 2021). Another study by the Dan team Reported that IgG-Abs against the Receptor binding domain have a shorter life than the Spike protein-Abs and were detectable only after three months maximum (Rydyznski Moderbacher et al.,

2020). Tan et al. suggest in their study that very severe manifestations accompany a sustained immune response, as these patients had high levels of IgG and IgM Antibodies (Wenting Tan et al., 2020). Zhang et al. team also noticed that in some infected pediatrics, the humoral response was detected by the measurable neutralizing antibodies against both the N-protein and the RBD (Y. Zhang et al., 2020). Zipeto et al team. Suggested that the RBD is hidden by protein folding till exact binding to ACE2, and virus neutralization can only achieved by antibody targeting this domain (Zipeto et al., 2020) (Li et al., 2022).

In this context, it is essential to mention that different studies' results may depend on the study's design basically and the targeted sample. The immune response in mild and moderate cases of SARS-Cov2 is similar to that in other viruses to a certain degree (Schulte-Schrepping et al., 2020). Still, in the severe COVID-19 illness, immunological studies showed some phenomenal lymphopenia, predominantly the decrease in the peripheral T cells (Chen & John Wherry, 2020), that may be due to their recruitment to the lymph node and their proliferation inhibition via the myeloid-derived suppressor-like cells (MDSC-like) which increased intensely in severe SARS CoV2 illness (Schulte-Schrepping et al., 2020). On the other hand, an increase in neutrophils is seen, accompanied by an increase in neutrophile precursors, indicating an emergency myelopoiesis (Schulte-Schrepping et al., 2020).

Also, another immune component increased in the severe cases of covid 19 is the Neutrophile extracellular trap (NET) formation; the disequilibrium in NET formation and degradation leads to an increase in the pathogenicity of the disease; these manifestations are associated with the risk of thrombotic formation conditions even when prophylactic anti-coagulation is giving (Zuo et al., 2021).

Also, an impairment of Natural killer function with increasing extended expression of INF-stimulated genes (ISGs) was seen in severe COVID-19. In contrast, the tumor necrosis factor (TNF) stimulated genes were dominant in mild and moderate disease forms (Benjam et al., 2020).

Also, the presence of C3a accompanied by severe illness was reported to bring an unusual T cell (CD16+), which has a unique role associated with releasing neutrophil and monocyte chemoattractants, This process is accompanied by microvascular endothelial cell injury that may be seen in multiple organ failure in fatal COVID-19 (Georg et al., 2021)

(Shafqat et al., 2022). Acute respiratory distress syndrome (ARDS) is the most fatal complication of COVID-19, caused by an increase in cytokines and chemoattractants during a cytokine storm. This leads to pneumonia, destruction of the lung epithelium, and multi-organ failure. Evidence suggests that an increase in IL-6 plays a critical role in inducing an excessive immune response, rather than the viral infection itself (Li et al., 2022) (Montazersaheb et al., 2022). A theory suggests that viral replication leads to pyroptosis (high inflammatory lytic apoptosis). This pyroptosis increases the release of the pro-inflammatory cytokines. Another theory suggests that binding neutralizing antibodies to the surface antigen induces an inflammatory reaction by recruiting macrophages and monocytes in the lungs via releasing IL-8 and monocyte chemoattractant protein (MCP)-1 (Montazersaheb et al., 2022). The specified adaptive immunity worked synchronously with the Antigen-presenting mechanism known in humans the Human Leukocyte antigen HLA.

2.5 Human Leukocyte Antigen System

The human leukocyte Antigen (HLA) system is the name of the Human Major Histocompatibility Complex (MHC); it was initially revealed in mice as a genetic locus accompanying the rejection of transplanted organs. In 1954, Rood and Dausset identified the same genetic system in a patient with multiple blood transfusions and called it the human leukocyte antigens. The polymorphisms and allele frequencies in the HLA system vary significantly across different populations based on ethnic and geographic factors. Certain populations may possess unique HLA alleles that occur at high frequencies among their individuals. This is important to consider in immunological studies such as vaccine development (Medhasi & Chantratita, 2022).

The HLA genes cluster located on chromosome number sex 6 (6p21.1 – 21.3), inherited by Mendel's Law, expressed in a co-dominance pattern, the sum of inherited HLA molecules form the individual haplotype pattern (Choo, 2007) (Medhasi & Chantratita, 2022). The HLA system is the most polymorphic in the human genome; it consists of three classes: Class I includes the genes coding three classic molecules, HLA-A, HLA-B, and HLA-C, and other genes coding less known non-classical HLA molecules, HLA-D, HLA-E, HLA-G, and HLA-F. Class II includes the genes coding HLA-DR, HLA-DQ, and HLA–DP, whereas class III

region includes genes responsible for coding immune regulatory proteins, genes coding proteins for the complement system, and TNF family genes (Medhasi & Chantratita, 2022).

Class I cluster is the closest to the telomere of the short arm. The genes within this region encode for the class I molecule α-polypeptide chain. Class I HLA molecules comprise three non-covalently linked parts: A heavy chain, a light chain, and a self-peptide. The heavy chain is a glycosylated amino acid. It is about 340 AA long, containing about thirty cytoplasmic amino acid residues and forty transmembrane amino acid residues. It has three extracellular domains (α1, α2, and α3), each is about ninety amino acid residues, α1 and α2 form the peptide-binding groove; this binding groove is classified into six discrete pockets (A–F) according to their chemical and physical prosperities (Choo, 2007) (Medhasi & Chantratita, 2022) (Szeto et al., 2021). Each extracellular domain in the heavy chain is encoded by different exons. Exon one encodes the leader peptide, while Exons two, three, and four are code α1, α2, and α3 domains, respectively. Exon five codes the transmembrane part, Exons six and seven code the cytoplasmic part, and Exon 8 codes the 3' UTR.

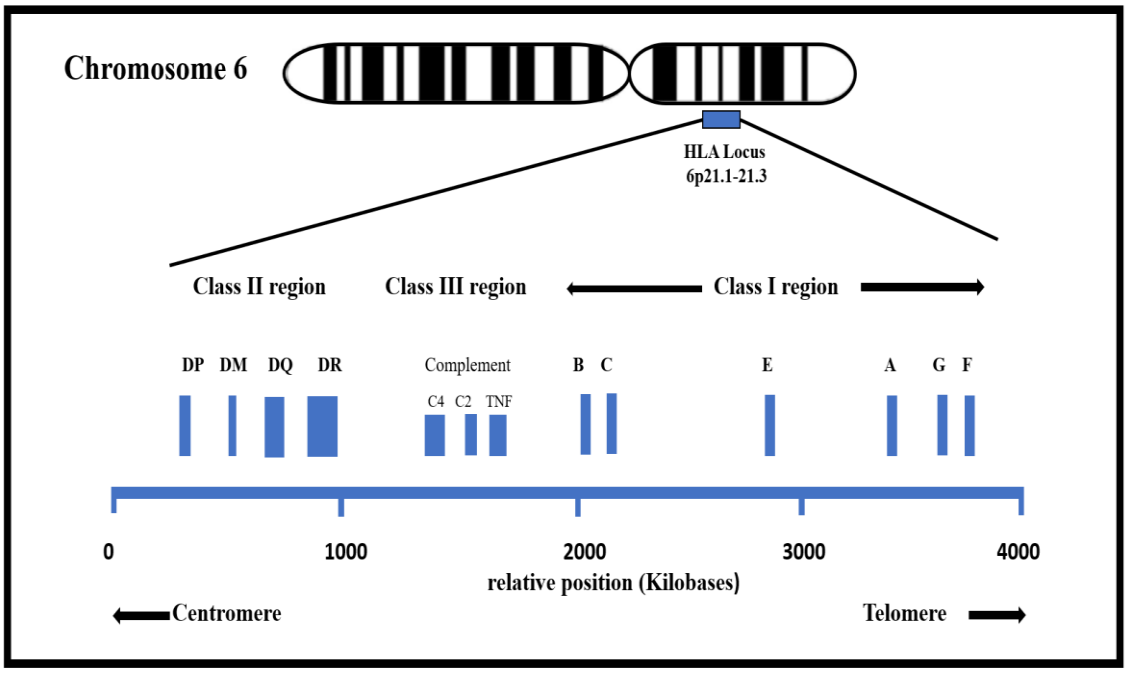

Figure 2.2 Schematic diagram of HLA Genes Locus on short the arm of chromosome six: showing the three classes of HLA system.
(https://onlinelibrary.wiley.com/doi/epdf/10.1111/iji.12513) (Arnaiz-Villena et al., 2021)

The β2-microglobulin protein forms the class I HLA light chain, encoded by the β2M gene on chromosome 15. It is linked to the α2 and α3 domains by non-covalent interactions. β2-microglobulin is necessary for stabilizing the HLA-I peptide complex and is responsible for the proper conformation of the binding groove (Medhasi & Chantratita, 2022). Once the heavy chain peptides are synthesized, an early folding event occurs within it, and a disulfide bond is formed in the presence of chaperone calnexin (CNX). After that, the heavy chain is linked to the β2-microglobulin and assembly in the endoplasmic reticulum (ER) as a heterodimer (Medhasi & Chantratita, 2022).

Antigenic proteins are processed by aminopeptidase ERAP1 and ERAP2 to shorter ones, about eight to ten amino acids (Seliger et al., 2006). These short peptides can be loaded to the HLA heterodimer that has been loaded into the peptide-loading complex (PLC); peptides binding to the HLA heterodimer according to their affinity to form the MHC-I peptide complex, the free amino acid carboxyl-terminus of these peptides associated with H-bonds to the α chain residues of the α chain (Szeto et al., 2021).

As mentioned earlier, α1 and α2 form the peptide-binding groove; this binding groove is classified into six discrete pockets (A–F) according to their chemical and physical prosperities (Szeto et al., 2021). The B and F pockets in the peptide-binding groove of class I HLA are the most important as they are the main anchor for peptide residue binding. Peptides that bind to the same HLA alleles frequently have identical conserved residues (Szeto et al., 2021). Then, these completely assembled complexes are transported to the cell surface via the Golgi apparatus after de-attaching from the PLC so they can present the peptides to the immune system.

Class II HLA molecules are encoded by the HLA genes region located toward the centromere; they encode two chains α and β, and each chain has two domains α1 and α2, β1 and β2, respectively. The leader peptide of both α and β chains of class II HLA is encoded by exon one, two extracellular domains, a transmembrane domain, and cytoplasmic tail encoded by exons (two and three), exon four, and exon five, respectively, in the β chain. While for the α chain, exon 4 codes both transmembrane and cytoplasmic domains (Medhasi & Chantratita, 2022) (Szeto et al., 2021).

Class II HLA molecules are synthesized in the endoplasmic reticulum and linked to an invariant chain. Class II HLA molecules have a cleft of 13–25 amino acids consisting of two

α-helixes walls and a β-sheet occupied by a small fragment of a consistent chain called a clip. The invariant chain stabilizes the binding cleft between the α1 and β1. It directs the molecules to the Golgi apparatus, preventing the binding of any peptides to the cleft until the HLA molecules fuse in endosomes containing an endocytosed peptide. The CLIP cleavage event occurred by the cathepsins, making the binding groove empty to be loaded by antigenic peptides. Then, these HLA-II-peptide complexes are transported to the cell surfaces to present their linked peptides to the immune system.

Class I HLA exists in any nucleated cell and present endogenous peptides for the CD8+ cytotoxic T cell, which recognizes the epitope by a particular receptor known as T Cell Receptor (TCR), kills the infected cell by releasing a toxic mediator and produces memory cells, while class II HLA exists only in antigen presenting cell APC, such as (monocyte, macrophage, dendritic cells, and B cells). This class is responsible for presenting the present exogenous peptides for CD4+ T helper cell (Choo, 2007) (Alicia, 2020) (Caterina A.M. La Porta, 2020). T cell receptors need to recognize the MHC-peptide complex (pMHC) to activate the T cell's response so that pMHC binds to the antigen-binding site of the TCR. TCR is a heterodimer composed of two subunits, either the α and β chains or the γ and δ chains. The most expressed is the αβ TCR; each subunit has two extracellular domains. A disulfide bridge binds between the two subunits of TCR (α and β), forming an immunoglobulin-like structure having an antigen binding site that is called the V domain. This V domain has three variable complementary determining regions (CDR1, CDR2, and CDR3). The CDR3 is the most variable one (Szeto et al., 2021) (F. Zhang et al., 2020) (Mariuzza et al., 2020) (Garcia & Adams, 2005).

The variety of TCRs relies on the philosophies of random segment arrangement of different genes mixed and the clonal selection; that is when TCR is activated toward specific antigenic peptides, this leads to the proliferation of more of this particular TCR cell clone (F. Zhang et al., 2020) (Garcia & Adams, 2005). This process activates the immune response to eliminate the antigen, however, in certain cases of covid 19, it may lead to an enormous life-threatening immune response as the fatal ARDS accompanied by severe infection (Gupta et al., 2020).

2.6 Human leukocyte Antigen alleles related to SARS CoV2

In COVID-19, the imbalance in immune response and the inflammatory process are responsible for severe symptoms such as acute respiratory distress syndrome (ARDS), secondary critical infections, cardiovascular disorders, deep venous thromboembolism, and organ failure (Nishiga et al., 2020). This is due to the triggering of proinflammatory cytokines and chemokines (IL-6, TNFα, and IL-1β), which attract more cytotoxic cells and neutrophils, potentially destroying surrounding tissues and causing severe systemic inflammation known as cytokine storm syndrome (Del Valle et al., 2020).

The HLA system plays a key role in infectious diseases as it is the primary mechanism for virus recognition. HLA loci exhibit 85% to 95% heterozygosity, providing a diverse capability to recognize different types of viruses (Alicia, 2020). During the SARS-CoV outbreak, several studies reported certain HLA alleles influencing the progression of the illness. Scientists believe that the genome similarity in SARS-CoV-2 may result in similar effects. A study dealing with the Italian population's HLA haplotypes supports the suggesting that population HLA polymorphisms are responsible for the different regional incidences and death cases of COVID-19 in Italy. They found that the two most frequent HLA haplotypes in the Italian population exhibited regional distributions that overlapped with COVID-19 incidence and death rates. The haplotype HLA-A*02:01,-B*18:01,-C*07:01,-DRB1*11:04, showed a negative association, while the haplotype HLA-A*01:01,-B*08:01,-C*07:01,-DRB1*03:01, showed a positive association, indicating increased susceptibility to the severity (Pisanti et al., 2020).

In another study among a cohort of 96 Northern Italian COVID-19 patients from Lombardy, the distribution of HLA alleles was analyzed to explore correlations with disease severity. The study findings linked the HLA-B*07 serotype to severe disease, while the HLA-B*27 and C*12:02 allele were associated with milder cases (Guerini et al., 2022). Despite lacking case-control examination, these studies highlight the impact of genetic background on COVID-19 severity. Another Italian study compared the HLA allele frequencies in a group of 99 severe COVID-19 patients with a control group of 1017 individuals and found significant correlations for HLA-DRB1*15:01, HLA-DQB1*06:02, and HLA-B*27:07 (Novelli et al., 2020). These studies predict different HLA alleles depending on the research methods used, but it remains important to consider that HLA allele frequencies are ethnic

and geographically dependent. Different populations may have different HLA alleles that react differently to the virus, explaining the pandemic's severity in certain countries.

China, where the virus first appeared, has conducted many studies to understand the genetic predisposition to COVID-19 infection and its severity. One study involving 82 COVID-19 patients revealed that alleles HLA-C*07:29, C*08:01G, B*15:27, B*40:06, DRB1*04:06, and DPB1*36:01 had higher frequencies in COVID-19 patients compared to the control group. Conversely, the frequencies of DRB1*12:02 and DPB1*04:01 alleles were lower among COVID-19 patients. After correcting for multiple testing, only alleles HLA-C*07:29 and HLA-B*15:27 remained statistically significant.(W. Wang et al., 2020) An other study of 332 Chinese patients indicated that the HLA-A*11:01, B*51:01, and C*14:02 alleles were significantly associated with severe disease prognosis. (F. Wang et al., 2020a)

Another study involving 159 Vietnamese patients and 52 uninfected controls explored HLA-I and HLA-II alleles, revealing significant variation in HLA allele frequencies between the two categories. The alleles HLA-A*03:01, HLA-A*30:01, HLA-DQA1*01:02, HLA-DRB1*15:01, and HLA-DRB5*02:02 were more frequent in the control group, while HLA-DRB1*09:01 was more common in patients. In the most severe patients, the alleles HLA-F*01:01, HLA-F*01:03, and DPA1*01:03 and DPA1*02:01, DPB1*04:01, DQA1*01:02, and DQB1*05:02 were associated with severity, whereas DOB*01:01, DRB1*05:01, and DRB1*09:01 were associated with milder disease. The findings of HLA-DQA1*01:02 and DRB1*09:01 alleles in mild and severe patients, respectively, and the control group were identified with potential dual roles in protection and predisposition (Nguyen et al., 2024).

In a study of 178 Japanese COVID-19 patients, researchers analyzed HLA-I and HLA-DRB1 regarding their association with severe COVID-19. Among 32 common HLA alleles, only HLA-DRB1*09:01 was strongly associated with COVID-19 by a significant statistical correlation with disease severity (Alitzel et al., 2021). This highlights the complex role of HLA class I and II alleles in COVID-19. Additionally, HLA-II alleles, particularly HLA-DR, play a key role in severe SARS-CoV-2 infections, with observed downregulation of their expression during critical illness (Spinetti et al., 2020) (Zmijewski & Pittet, 2020). This suggests a significant role in antigen presentation mechanisms, indicating that differences in HLA class II alleles can impact the immune response. Interestingly, the expression of HLA-

DR decreases in older individuals, indicating increased susceptibility to severe infection (Zmijewski & Pittet, 2020).

In the Middle East, a region familiar with MERS-CoV in 2012, a significant percentage of SARS-CoV-2 incidence was observed. An Iranian study of 48 ICU COVID-19 patients versus 500 controls revealed a significant relation between HLA-B*38, HLA-A*68, HLA-A*24, and HLA-DRB1*01 to severe COVID-19 to the predisposition to severe form of the Covid-19 illness. Notably, they Didn't perform any Exclusion criteria for other risk factors for severe forms of illness or perform ordinal logistic regression modeling (Farahani et al., 2021). Another recent Iranian cross-sectional study analyzed genomic DNA from 109 COVID-19 patients and 70 healthy controls, focusing on HLA-A, HLA-B, and HLA-DRB1 alleles. Results showed that HLA-DRB1*11:01 and HLA-DRB1*04:03 were more frequent in severe cases, while HLA-DRB1*04:01 was more common in moderate cases and healthy controls. Additionally, HLA-B*07:35 and HLA-DRB1*07:01 were higher in patients than controls. Conversely, HLA-B*51:01, HLA-DRB1*11:05, HLA-DRB1*13:05, and HLA-DRB1*14:01 were more frequent in healthy individuals, suggesting a potential protective role against COVID-19 (Abolnezhadian et al., 2024).

A study of 115 individuals in the UAE of Emirati nationality found that HLA-B*51:01 and A*26:01 alleles had a negative relation, while HLA-A*03:01, HLA-DRB1*15:01, and the B*44 supertype had a positive correlation with illness severity (Alnaqbi et al., 2022). Another UAE study with a cohort of 92 individuals of different nationalities found that HLA-C*04 and HLA-B*35 alleles were significantly associated with COVID-19 severity. The class I haplotype HLA-C*04-B*35 showed a strong correlation with severe disease outcomes. Notably, HLA-C*04 and HLA-B*35 have been linked to other respiratory conditions, such as pulmonary arterial hypertension (Tay et al., 2023).

In Egypt, a study of 69 moderate and severe COVID-19 patients indicated that alleles HLA-*B41, HLA-B*42, HLA-C*16, and HLA-C*17 were associated with COVID-19 severity, whereas HLA-B*15, HLA-C*07, and HLA-C*12 were associated with protection. After applying ordinal regression analysis, only HLA-B*15 showed a significant correlation with protection (Abdelhafiz et al., 2022). In Spanish Mediterranean Caucasian patients with moderate and severe illness, the analysis of a cohort of 72 individuals indicated that HLA-C*08:02, HLA-C*12:03, and HLA-C*16:01 were more common in mild cases than severe

ones, supported by their affinity for binding a greater number of SARS-CoV-2 epitopes (Vigón et al., 2022).

On the other hand, many bioinformatic analyses are concerned with viral epitope-HLA binding affinity. Predicting epitopes of the SARS-CoV-2 viral structural protein, mainly the S protein, and sometimes other proteins such as the N protein, involves determining binding thresholds, which are crucial for these analyses (Barquera et al., 2020) (Iturrieta-zuazo et al., 2020a) (Nguyen et al., 2020) (F. Wang et al., 2020b). A study conducted on 438 HLA-I alleles and the proteomes of seven pandemic viruses, including HIV, Influenza, and coronaviruses, classified the affinities into four different categories. Among the strongest HLA binders, 40 alleles were identified. Specifically, HLA-A*02:02, B*15:03, DRB1*01:02, A*68:01, B*15:25, C*03:02, DRB1*07:01, A*02:01, and C*14:02 were found to bind to coronaviruses (Barquera et al., 2020).

In this context, Nguyen et al. conducted an extensive in-silico analysis examining the binding affinity of 145 HLA-A, HLA-B, and HLA-C genotypes regarding SARS-CoV-2 peptides and potential cross-protective immunity from prior exposure to other coronaviruses (HKU1, OC43, NL63, and 229E). They found that the alleles HLA-A*02:02, HLA-B*15:03, and HLA-C*12:03 had the highest capacity for presenting conserved peptides, while A*25:01, B*46:01, and C*01:02 had the lowest capacity. Notably, HLA-B*46:01 had the fewest predicted binding peptides for SARS-CoV-2, indicating that individuals with this allele might be particularly susceptible to COVID-19 (Nguyen et al., 2020). Based on Nguyen's approach, a Spanish study on 45 COVID-19 patients observed theoretically protective alleles, such as HLA-B*15:03, in patients with severe disease who died, and alleles with low affinity for the virus, such as HLA-A*25:01, in patients with moderate disease progression. Additionally, the frequency of some of these alleles, such as HLA-B*46:01, was almost nil among the Spanish population.(Iturrieta-zuazo et al., 2020a) Similarly, a Russian study, also utilizing Nguyen's approach, developed a risk score to evaluate the ability of HLA Class I to present SARS-CoV-2 peptides. This study involved a cohort of 111 deceased patients and 428 volunteers. The deceased patients were divided into two groups: 26 adults under 60 years old and 85 adults over 60 years old. Using NGS for HLA-I genotyping, the resulting risk score was significantly higher in the group of deceased adults compared to elderly adults. Notably, the presence of the HLA-A*01:01 allele was associated with high

risk, while HLA-A*02:01 and HLA-A*03:01 were primarily linked to low risk. Homozygosity in patients strongly reinforced these findings (Shkurnikov et al., 2021)

A Brazilian study used molecular dynamics to predict the best-fitted S protein epitopes. They identified 24 immunogenic epitopes that could interact with 17 different MHC-I alleles (HLA- A*01:01; HLA- A*02:01; HLA- A*11:01; HLA- A*24:02; HLA- A*68:01; HLA-A*23:01; HLA- A*26:01; HLA- A*30:02; HLA- A*31:01; HLA- B*07:02; HLA- B*51:01; HLA- B*35:01; HLA- B*44:02; HLA- B*35:03; HLA- C*05:01; HLA- C*07:01 and HLA-C*15:02) which are recognized at high frequencies in the Brazilian population (De Moura et al., 2020). A vaccine development study tried to identify potential vaccine targets aligned with the epitopes tested for both T cell activation and MHC binding across the CoV-2 genome sequences. They identified 27 epitopes identical to segments of the S and N CoV2 protein and then analyzed the frequencies of related HLA. This emphasis on HLA studies is further supported by another study among 1351 Italian subjects who received the Pfizer-BioNTech vaccine. They analyzed antibody levels post-vaccination and HLA genotypes, revealing a strong association with 12 variants (HLA-A*03:01, HLA-C*12:02, HLA-B*52:01, HLA-A*29:02, HLA -DQB1*06:01, HLA-DRB1*15:02, HLA-DQB1*02:01, HLA-DRB1*14:01, HLA-DQA1*01:01, HLA- DQA1*02:01, HLA-DQB1*05:03, and HLA-DRB1*07:01). Specifically, the HLA-A*03:01 allele was most significantly associated with serum IgG levels (Esposito et al., 2024).

A recent Brazilian study resembling this analysis examined antibody variability in vaccine response among a healthy cohort divided into two groups: low and high-stable antibody individuals. This study indicated the significance of differential splicing regulatory mechanisms, mainly concerning HLA alleles, in delineating vaccine immunogenicity. The study revealed that HLA-A*02:01, HLA-DQB1*06:02, HLA-DRB1*07:01, HLA-B*40:01, and the haplotype HLA-DRB1*15:01-DQA1*01:02-DQB1*06:02 were related to a high COVID-19 vaccine response. Conversely, HLA-A*03:01, HLA-B*08:01, HLA-B*18:01, HLA-C*07:01, and the haplotype HLA-DRB1*01:01-DQA1*01:01-DQB1*05:01 were related to a low vaccine response (Santos-Reboucas et al., 2024).A study supporting the findings of bioinformatic predictions tested HLA-A*02:01-restricted SARS-CoV-2 T-cell epitopes predicted by immunoinformatic approaches and found that they can react to

antibodies in the blood of covid-19 patients which means they are indeed immunogenic in natural SARS-CoV-2 infections of HLA-A*02:01 patients (Tan et al., 2024).

A recent American study investigated the role of classical HLA genotypes in asymptomatic COVID-19. They particularly examined HLA-B*15:01, previously suggested as a protective factor associated with asymptomatic illness due to its reaction to the NQKLIANQF epitope conserved in seasonal coronaviruses OC43-CoV and HKU1-CoV (Augusto et al., 2022). However, they found no association with the asymptomatic form and identified three patients with severe COVID-19 (Marchal et al., 2024). Additionally, a genome-wide association study (GWAS) analyzed the association between COVID-19 severity and HLAs in 435 COVID-19 patients from Switzerland, the United States, Spain, and Germany. This study indicated a potential double risk of intubation in patients with HLA-C*04:01 due to fewer predicted SARS-CoV-2 epitopes (Weiner et al., 2021).

Based on these observations, it can be concluded that contradictory findings arise from various potential reasons, including (1) Genetic Diversity: Different populations have unique HLA allele distributions, which can lead to varying results. What is significant in one population may not be in another due to different genetic backgrounds. (2) Sample Sizes and Study Designs: Variations in sample sizes, study designs, and methodologies can lead to different conclusions. Larger, well-designed studies tend to provide more reliable results. (3) The environmental factors: differences in environmental factors, healthcare systems, and population health can influence the outcomes of these studies. (4) Reporting and Publication rates: Studies with significant findings more commonly to be published, while those with insignificant results might not be, tilting the overall picture.

Chapter Three: Methodology

A correlational study was conducted to statistically evaluate a possible correlation between HLA-I and Covid 19 severity among the Palestinian population. Advanced and robust analysis techniques in genotyping and sequencing were employed, with ethical approval obtained to ensure adherence to ethical standards. The analysis was carried out with integrity and objectivity, avoiding any bias in the interpretation of the results.

3.1 Study Design

The research methodology was precisely designed to ensure the accuracy and validity of the target sample. This involved a comprehensive literature review to identify and evaluate various methods used in previous studies concerning the associations between HLA-I alleles and disease severity. The outlined steps are shown in the schematic below figure (3.1)

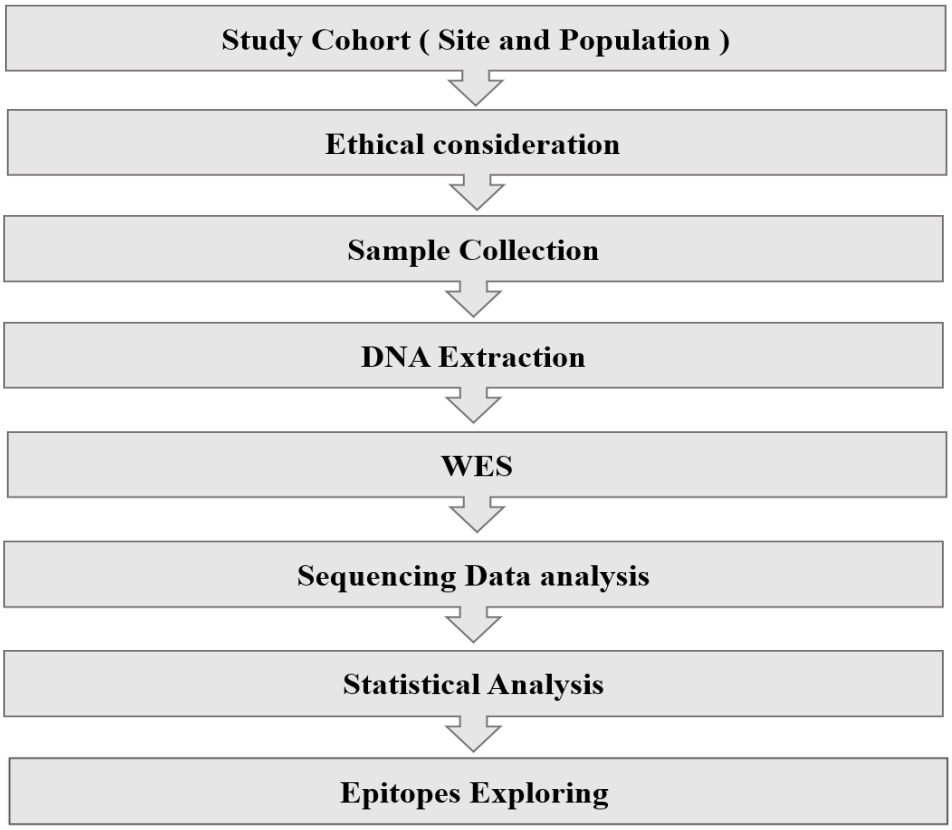

Figure 3.1 Schematic of the Study Design

3.2 Study Cohort (Site and sample)

The sample population was 18 critical Palestinian (COVID-19) patients: 3 patients on a high flow oxygen that is 40 L/min or more, and 15 were admitted to the ICU, where 7 of them died. All patients were under the age of 55 of either gender with no documented history of any pre-existing diseases, chronic conditions, or significant health-related abnormalities. Samples and data were collected retrospectively from the governmental hospitals. The research was conducted in governmental hospitals Jenin Governmental Hospital in Jenin, Al Hilal Hospital and Alaskary Hospital in Nablus, Hugochavies Hospital and Palestinian Medical Complex in Ramallah, Alia, Dora, and Yatta hospitals in Hebron, and southern west bank - occupied Palestinian territories. The Distribution of patients is illustrated in Appendix A.

3.3 Ethical consideration

Blood samples and clinical data for the patients were retrospectively provided by the governmental hospitals under the approval and facilities provided by the Palestinian Ministry of Health. samples were assigned identification codes to keep Patients' confidentiality. The Helsinki Committee for Ethical Approval granted its approval, with the IRB number being RHRC/HC/872/21, attached as appendix F.

3.4 Sample collection

EDTA whole blood samples were collected between February 2021 and August 2021, each patient provided approximately 3 mL of blood, with strict precautions taken to prevent infection. Samples were preserved in hospital refrigerators and then transported in cooled preservative containers to maintain their validity. Subsequently, the buffy coat was separated and stored at -70°C in a freezer until DNA extraction was performed.

Prior studies investigating the correlation between HLA-I and COVID-19 severity have yielded conflicting results, highlighting the importance of study design and sample selection. Given the ethnic and geographical specificity of HLA genes, this research aims to account for these factors carefully. Moreover, in light of the diverse immune responses observed in COVID-19 patients, specific inclusion and exclusion criteria have been established to ensure

a more accurate analysis. The inclusion criteria for the selected samples are outlined in figure (3.2). Samples were coded from IC01 to IC56, and final clinical data were acquired retrospectively

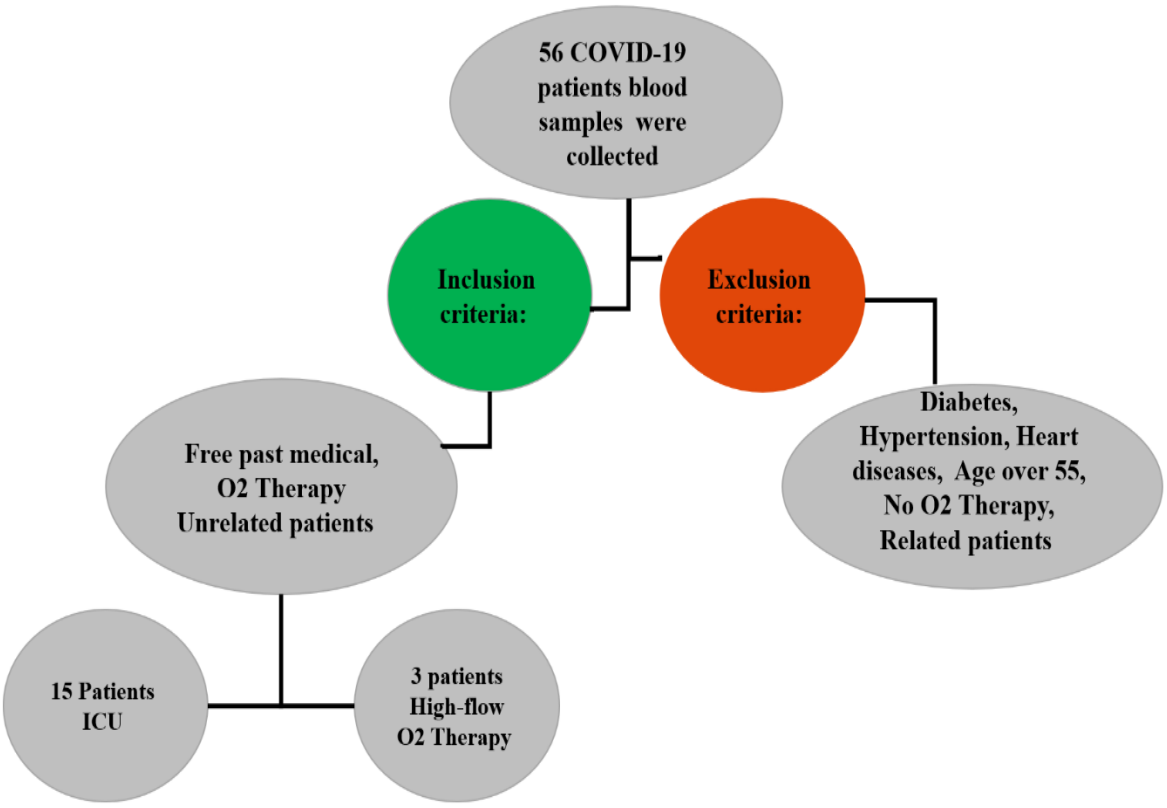

Figure 3.2 : The Inclusion and exclusion Criteria for the study sample.

3.5 HLA Genotyping

3.5.1 DNA Extraction

The DNA extraction was carried out utilizing the Wizard® Genomic DNA Purification Kit from Promega, following the prescribed protocols for Genomic DNA Isolation from Whole Blood (www.promega.com)). Initially, 300 µL of the buffy coat was gently mixed with 900 µL of Cell Lysis Solution in a sterile 1.5 mL microcentrifuge tube. Following a 10-minute incubation at room temperature, centrifugation at 13,000–16,000 × g for 20 seconds effectively separated the cellular components, preserving an undisturbed white pellet.

The subsequent step involved resuspending the white blood cells, adding 300 µL of Nuclei Lysis Solution, then adding 100 µL of Protein Precipitation Solution. After vigorous mixing by the vortex and centrifugation at 13,000–16,000 × g for 3 minutes, the supernatant was carefully transferred to 300 µL of isopropanol, mixed thoroughly, and centrifuged again. The resultant pellet was washed twice with 70% ethanol, followed by resuspension in 100 µL of DNA Rehydration Solution.

Evaluation of the DNA integrity was conducted through Agarose Gel Electrophoresis figure (3.3), while a Nanodrop spectrophotometer was employed to determine concentration and purity, based on absorbance ratios at 260/280 and 260/230.

Concentrations and 260\280 and 260\230 ratios are illustrated in Appendix B.

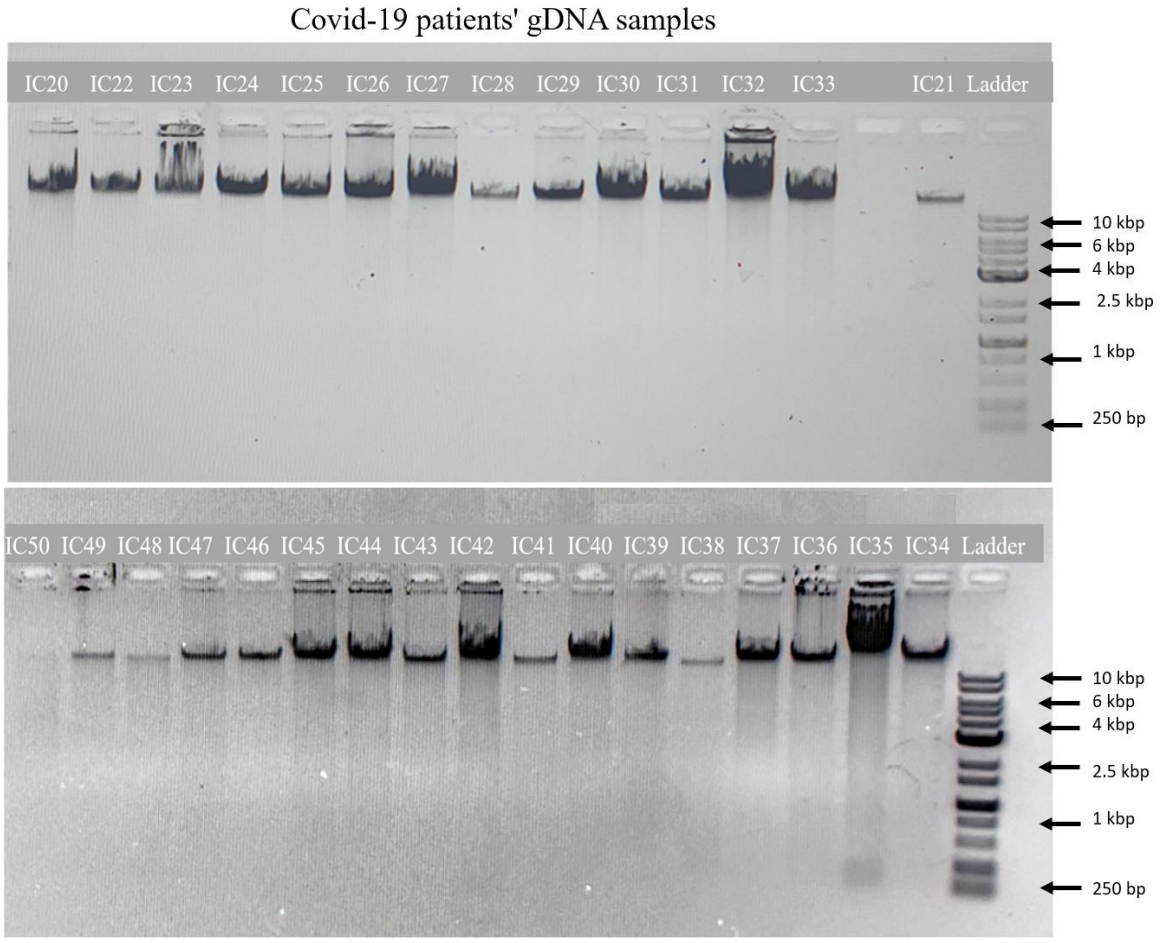

Figure 3.3 The Agarose Gel Electrophoresis image of the extracted DNA.

The figure illustrates the intact nature of the extracted DNA, showing no evidence of fragmentation. For each sample, a mixture was prepared consisting of 5 µL of extracted

genomic DNA, 3 µL of loading buffer, and 7 µL of H2O. These mixtures were then loaded onto a 75% agarose gel and subjected to electrophoresis for 30 minutes at 400 volts and 100 amperes.

3.5.2 Next Generation Sequencing

Next Generation Sequencing (NGS) was used to identify HLA genotyping with high accuracy and 4-digit precision by applying whole exome sequencing, NGS has three main steps as in figure (3.4). For library preparation and enrichment, an initial gDNA concentration of 500 ng in 30 µL was prepared for each sample, as illustrated in Appendix B, following the workflow steps outlined in the Illumina DNA Prep with Enrichment Reference Guide. (https://support.illumina.com/downloads/illumina-dna-prep-with-enrichment-reference-guide-1000000048041.html).

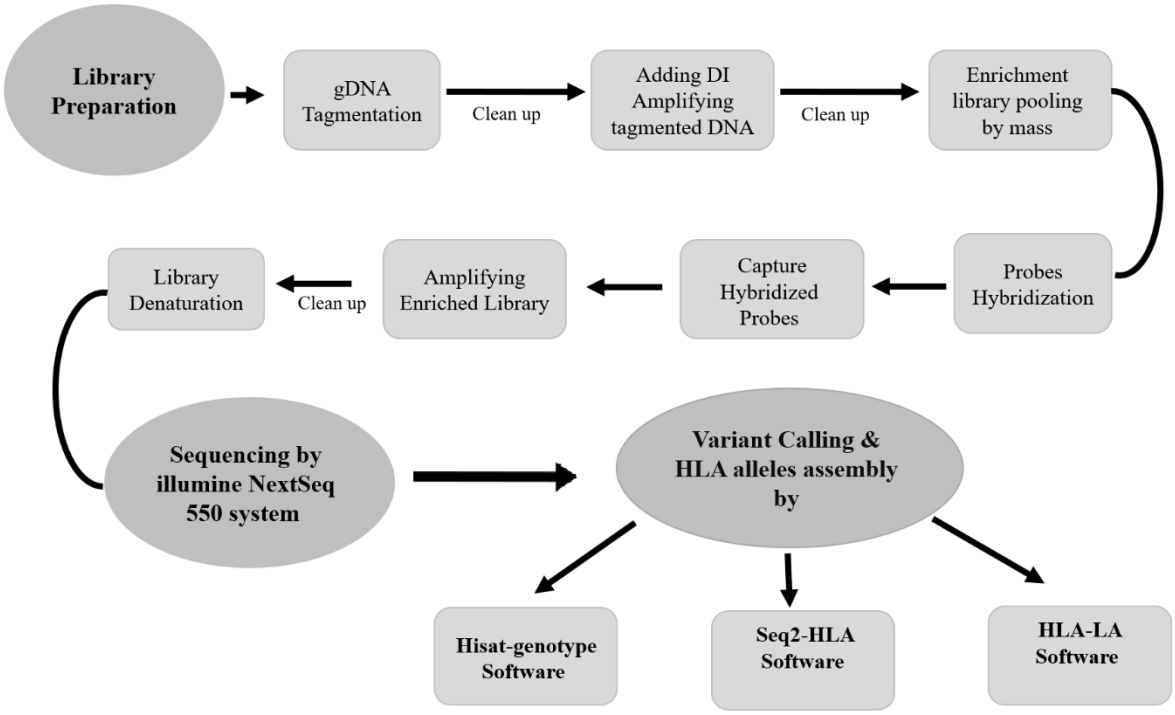

Figure 3.3 : NGS Data Analysis Workflow.

It consists of three main steps: library preparation, sequencing using the Illumina NextSeq 550, and data analysis to perform variant calling. The primary steps of library preparation are outlined, library denaturation with 0.2N NaOH, followed by dilution with HT1 before

loading into the Illumina NextSeq 550 cartridge. Data analysis is conducted using three different algorithm software programs: HISAT-genotype, HLA-LA, and Seq2HLA, all of which typically perform variant calling as part of their analysis pipeline.

First, tagmentation of gDNA was performed using eBLT (enrichment-based linked transposons), adding sequences to the gDNA that act as priming sites. Next, a post-tagmentation clean-up was conducted using TWB (Tagment Wash Buffer) to stop the tagmentation reaction and wash the tagmented DNA on eBLT. A dual unique index was then added, and the tagmented DNA was amplified using a thermal cycler, as illustrated in figure (2.5). The library was cleaned up by size selection using AMPure XP beads in two sequential washes: the first to eliminate large DNA fragments and the second to remove small fragments. For enrichment library pooling, equal quantities of each sample were ensured by pooling 12-plex together for each run. Concentrations were measured using a Quantus™ fluorometer from Promega to achieve a final library volume of 30 µL.

	1	2	3	4	5	6	7	8	9	10	11	12
A			IC1	IC20	IC32	IC53						
B					IC34	IC54						
C				IC22	IC36	IC55						
D				IC23	IC14	IC56						
E				IC24								
F				IC49								
G			IC7	IC26								
H				IC30	IC51							

Figure 3.4 : A schematic diagram illustrating the distribution of study samples.
For the dual unique index employed in library preparation and enrichment. The order and codes assigned to the samples have been meticulously reviewed and approved, confirming their inclusion in the research study.

Probes were hybridized by combining the pre-enriched library DNA with NHB2 and the panel, followed by the addition of EHB2 buffer. This process facilitated the binding of targeted regions of the pre-enriched library DNA to specific capture probes. The hybridized probes were immediately captured using Streptavidin Magnetic Beads (SMB beads), subjected to successive washes with EEW, and eluted with EE1 buffer and HP3.

After elution, the enriched library underwent amplification using an appropriate amplification cycle tailored to the panel. The process included an initial denaturation at 95°C for 13 minutes and 98°C for 2 minutes, followed by 21 cycles at 98°C for 15 seconds and 60°C for 2 minutes, with a final extension at 72°C for 5 minutes. At the end of the program, samples were maintained at 4°C. The enriched library was subsequently cleaned using AMPure XP beads and eluted in RSB buffer. During the Cartridge loading step, PhiX Control v3 was included alongside the library. This control, a pre-prepared and consistent adapter-ligated library, served as a quality monitoring control for Illumina sequencing runs.

The second main step in NGS, sequencing, was conducted using the Illumina® NextSeq™ 550 system, which provides high throughput. The reagent cartridge and flow cell were prepared, and the manufacturer's software instructions were applied to set up and initiate the run. Cluster generation and sequencing were performed on the instrument, which automatically initiated an instrument wash using components already loaded on the system at the conclusion of the run.

During cluster generation, single DNA molecules bound to the surface of the flow cell were amplified to form clusters. These clusters were imaged using two-channel sequencing chemistry and specific filter combinations for each fluorescently labeled chain terminator. Once imaging of a tile on the flow cell was completed, the next tile was imaged, with this process repeated for each cycle of sequencing. The software handled base calling, filtering, and quality scoring as part of the image analysis. Run progress and statistics were accessible from selected locations.

As the run progressed, the control software automatically transferred base call (BCL) files to a designated output location, such as BaseSpace Sequence Hub or Local Run Manager, for secondary analysis. The BCL files were converted to FASTQ format using bcl2fastq, which demultiplexed the reads and trimmed adapter sequences. These FASTQ files were subsequently used for the final key step, variant calling, via HLA typing software.

The resulting FASTQ files were then processed using three HLA typing tools: HLA-LA, HISAT-genotype, and Seq2HLA. These different analysis programs, each based on distinct algorithms. HLA-LA is an alignment-based program that compares linear sequenced HLA regions with reference haplotypes, achieving approximately 93% accuracy for whole exome sequencing (WES). This program required FASTQ input, processed the reads to align them with reference sequences, and performed variant calling to identify HLA alleles (Dilthey et al., 2019).

HISAT-genotype, the second program, used two tools, hisatgenotype_extract_reads.py and hisatgenotype.py, with greater than 90% accuracy at read depths over 100X and up to 97% accuracy for the first and second fields in WES. This tool aligned sequencing reads to a reference genome and performed variant calling to determine genotypes at various loci (Liu et al., 2021) (Thuesen et al., 2022). The third program, Seq2HLA, conducted HLA typing from RNA-Seq data using an alignment approach. It provided four-digit resolution with accuracy exceeding 95% for the first field and approximately 75% for the second field. (Liu et al., 2021) This software required raw RNA-Seq reads in FASTQ format, aligned them to an HLA reference database, and performed variant calling.

All programs were accessed and downloaded from GitHub (https://github.com).

3.6 Statistical Analysis

For HLA alleles frequency calculation and comparison, first, the Allele Frequency Net Database (http://www.allelefrequencies.net/) was used, considering the Israel Arab pop 2 as a control group (Manor et al., 2016) for each HLA-A and HLA-B alleles while considering the HLA of Jordanians as a control for the HLA-C allele (Sánchez-Velasco et al., 2001); as the first (Israel Arab pop 2) control group does not have genotype for the HLA-C alleles.

- HLA allele frequencies were calculated by the following equation for both the study sample and the control group:

$$\text{HLA allele frequency} = \frac{\text{Number of specified allele repetition}}{2 * \text{Number of the sample individuals}} * 100\%$$

- Using Prism 5. Program each of Odd's ratios, 95% confidence interval, and P-values were calculated by choosing Fisher's exact test and Yates continuity-corrected $\chi 2$. Considering the

repetition of a specified allele is the number of successes, the number of alleles that failed is calculated by subtracting the number of successes from the total number of alleles for each group.

The final study sample is composed of 18 individuals for HLA alleles subclasses, while the control group (Israel Arab pop 2**)** consists of 12301 individuals sourced from the Ezer Mizion Bone Marrow Donor Registry in Israel. This choice was influenced by the significant number of Palestinians receiving transplantation treatment in Israeli hospitals, along with the historical and familial connections between the Israel Arab population and Arabs in the West Bank, characterized by cross-marriages and shared ancestry (Manor et al., 2016) and the control group for HLA-C consists of 146 individuals. This group was considered appropriate for the study, as clarified by the study that notably, approximately half of Jordanians have Palestinian origins, substantial number of Palestinian Arabs, around 1.5 million, are registered as refugees and displaced persons in Jordan, with many holding Jordanian citizenship. The connection between the Jordanian and Palestinian populations, along with shared ancestry, made this cohort a relevant choice for the investigation (Sánchez-Velasco et al., 2001). This huge difference in sample sizes made it crucial to do P-value correction. This was done by the Benjamini-Hochberg method.

- The Benjamini-Hochberg equation was obtained to calculate the Corrected P-values as the following formula:

$$\text{Corrected P-value} = \text{P-value} * \frac{\text{Number of tested P values}}{\text{Rank of p-value}}$$

A significant allele is the one that has a corrected P value <0.05

3.7 Epitopes exploring.

The Immune Epitope Database (IEDB) was explored to uncover epitopes associated with the significant HLA Class I alleles identified in the study concerning the spike glycoprotein and nucleocapsid of SARS-CoV-2. All identified epitopes for the spike glycoprotein have

been documented and are provided in Appendix C while for the nucleocapsid epitopes for the significant alleles are listed in Appendix D.

Utilizing the specified filter criteria: Epitope: any. Epitope Source: SARS-CoV2 (ID:2697049) SARS2, Spike glycoprotein [P0DTC2] (SARS-CoV2). Host: Human. Assay: Positive T cell response. MHC Restriction: Specify the desired HLA allele.

Subsequently, Python code was developed to visualize these epitopes' location concerning the spike protein of SARS-COV-2. This aimed to investigate the potential presence of a shared and specifically targeted domain.

The Python code link and data set link are attached in the Appendix section.

Chapter Four: Results

The study aimed to identify a potential correlation between HLA-I genotypes and the severity of COVID-19 among the Palestinian population with a history of severe and critical forms of the disease. Analysis of the enrolled patients' Whole Exome Sequencing (WES) data identified their HLA-I alleles and subsequent statistical analysis suggested the significance of these findings.

4.1 Genotyping Class I HLA using three distinct software programs.

To perform HLA-I genotyping of the enrolled samples, Whole Exome Sequencing (WES) was conducted using the Illumina® NextSeq™ 550 system, which provides high throughput. The outcomes, generated as BCL files in FASTQ format, were subsequently analyzed for variant calling. These variant calling files were processed using three distinct software programs, each employing different algorithms: HLA-LA, Hisat-genotype, and Seq2HLA.

The samples were processed in two runs. The first run included thirteen patients initially analyzed using HLA-LA, which employs alignments of linear sequenced HLA in BCL files with reference haplotypes, focusing on genomic differences in coding exons' antigen recognition domains. HLA-LA has a reported accuracy of 93% for WES, though this decreases due to potential miscalls in single nucleotides. To validate the genotyping, these samples were re-analyzed using Hisat-genotype, which utilizes C++ and Python programs (hisatgenotype.py and hisatgenotype_extract_reads.py) and claims over 90% accuracy with read depths above 100X, potentially reaching 97% accuracy for first and second fields in WES (Thuesen et al., 2022) (Liu et al., 2021).

A third program, Seq2HLA, was also used to cross-validate the results of the first run. Seq2HLA, which uses an alignment-based approach to provide four-digit resolution, reports over 95% accuracy for the first field and approximately 75% for the second field. The results from Seq2HLA were consistent with those obtained using Hisat-genotype. Based on this evaluation, Hisat-genotype was adopted for its higher accuracy. Consequently, the second run, which included five samples, was analyzed exclusively using Hisat-genotype, as shown in Table (4.1). This multi-program analysis approach ensured the reliability of the genotyping

results by identifying and resolving discrepancies between the tools. Differences marked in red font.

Table 4.1 : The HLA alleles for COVID-19 patients were obtained by three different software.

Patient Code	HLA-LA		Hisat-genotype		Seq2HLA	
IC 34	A*23:01	A*03:01	A*23:01	A*32:01	A*24:02	A*32:01
IC 7	A*24:02	A*34:02	A*34:02	A*24:02	A*34:02	A*24:02
IC 36	A*03:01	A*68:01	A*03:01	A*68:01	A*03:01	A*68:01
IC 14	A*03:01	A*03:01	A*01:01	A*32:01	A*01:01	A*32:01
IC 20	A*24:02	A*03:01	A*01:01	A*24:02	A*01:01	A*24:02
IC 22	A*03:01	A*03:01	A*32:01	A*02:01	A*32:01	A*02:01
IC 23	A*24:02	A*24:02	A*24:02	A*02:01	A*24:02	A*02:034
IC 1	A*03:05	A*30:02	A*01:01	A*30:02	A*01:01	A*30:02
IC 24	A*01:01	A*01:01	A*01:01	A*26:01	A*01:01	A*26:01
IC 49	A*03:01	A*03:01	A*29:02	A*02:01	A*29:02	A*02:01
IC 26	A*03:05	A*03:05	A*01:01	A*02:01	A*01:01	A*02:01
IC 30	A*30:01	A*03:01	A*30:01	A*03:01	A*30:01	A*03:01
IC 32	A*01:01	A*01:01	A*01:01	A*01:01	A*01:01	A*29:05
IC51			A*01:03	A*24:02		
IC53			A*29:01	A*01:01		
IC54			A*24:03	A*26:01		
IC55			A*30:02	A*33:01		
IC56			A*02:01	A*11:01		

Patient Code	HLA-LA		Hisat-genotype		Seq2HLA	
IC 34	B*07:02	B*07:02	B*49:01	B*49:01	B*49:01	B*49:01
IC 7	B*15:03	B*38:01	B*38:01	B*15:03	B*38:01	B*15:03
IC 36	B*35:01	B*35:01	B*08:01	B*35:08	B*08:01	B*35:08
IC 14	B*48:04	B*48:04	B*41:01	B*13:02	B*41:01	B*13:02
IC 20	B*35:01	B*40:06	B*35:01	B*40:06	B*35:08	B*40:06
IC 22	B*27:05	B*27:05	B*27:05	B*57:03	B*27:05	B*57:03
IC 23	B*40:01	B*78:02	B*40:01	B*52:01	B*40:01	B*52:01
IC 1	B*15:10	B*15:10	B*15:10	B*50:01	B*15:10	B*50:01
IC 24	B*35:01	B*38:01	B*38:01	B*35:01	B*38:01	B*35:01
IC 49	B*27:05	B*27:05	B*57:03	B*44:03	B*57:03	B*44:03
IC 26	B*57:01	B*57:01	B*57:01	B*49:01	B*57:01	B*49:01
IC 30	B*35:01	B*35:01	B*35:01	B*35:02	B*35:01	B*35:02
IC 32	B*37:01	B*51:01	B*37:01	B*51:01	B*37:01	B*51:01

IC51			B*44:03	B*35:02		
IC53			B*57:03	B*07:05		
IC54			B*41:02	B*51:01		
IC55			B*53:01	B*14:02		
IC56			B*08:01	B*51:01		

Patient Code	HLA-LA		Hisat-genotype		Seq2HLA	
IC 34	C*07:01	C*07:01	C*07:01	C*07:01	C*07:01	C*07:01
IC 7	C*12:03	C*12:03	C*12:03	C*03:04	C*12:03	C*03:03
IC 36	C*07:02	C*07:02	C*07:02	C*04:01	C*07:02	C*04:01
IC 14	C*07:04	C*07:04	C*07:04	C*16:04	C*07:04	C*07:04
IC 20	C*04:04	C*12:02	C*12:02	C*04:01	C*04:04	C*12:02
IC 22	C*07:01	C*02:02	C*07:01	C*02:02	C*07:01	C*02:02
IC 23	C*12:02	C*12:02	C*12:02	C*03:04	C*12:02	C*03:04
IC 1	C*06:02	C*06:02	C*06:02	C*03:04	C*06:02	C*03:03
IC 24	C*12:02	C*04:29	C*12:03	C*04:01	C*12:03	C*04:01
IC 49	C*07:01	C*07:01	C*07:01	C*16:01	C*07:01	C*16:01
IC 26	C*07:02	C*07:02	C*06:02	C*07:01	C*06:02	C*07:01
IC 30	C*07:02	C*07:02	C*04:01	C*07:01	C*04:04	C*07:01
IC 32	C*06:04	C*15:02	C*06:02	C*15:02	C*06:02	C*15:02
IC51			C*07:06	C*04:01		
IC53			C*07:01	C*15:05		
IC54			C*17:03	C*14:02		
IC55			C*06:02	C*08:02		
IC56			C*07:02	C*14:02		

Alleles marked in red indicate discrepancies between the three different algorithms.

4.2 The frequency and odds ratio values of the HLA alleles with uncorrected significance in COVID-19 patients.

To identify significant HLA alleles and their role in COVID-19 severity, allele frequencies in the study sample were calculated and compared with those in the control group (HLA allele frequencies in the Palestinian population). Notably, the existing literature on HLA alleles related to SARS-CoV-2 often does not specify a particular software program for NGS variant calling. Therefore, the results from the three different programs were used in the analysis to ensure cross-validation and to highlight the differences in allele significance based on the accuracy of the genotyping software. Statistical analysis was performed using the Prism program V5 to calculate the p-value, 95% confidence interval (CI), and odds ratio

for each allele identified in the study by applying Fisher's exact test and Yates continuity-corrected χ^2.

The control group size for HLA-A and HLA-B comprised 12,301 individuals, sourced from the Allele Frequency Net Database (http://www.allelefrequencies.net/). This population represented the Israel Arab population 2, as specified in the methodology chapter. Meanwhile, the control sample size for HLA-C was 146 individuals, utilizing the HLA data of Jordanians as the control for HLA-C alleles from the same database website, as the first control group does not have HLA-C genotyping. The significant alleles are those having a p-value <0.05, Odds ratio >1 indicates a positive association with increased SARS-CoV-2 severity; if it is <1, it indicates a negative association, meaning it may be considered a protective agent. The Confidence Interval 95% provides the uncertainty of the odds ratio value, this indicates a 95% confidence that the true odds ratio lies within the specified confidence interval, providing a reliable basis for assessing whether an allele acts as a protective or promoting factor. However, these alleles with a p-value <0.05 need to be corrected due to the huge difference in sample and control group sizes.

Initial data analysis suggested different significant alleles based on results from different softwire programs. Alleles with a p-value <0.05 in Hisat-genotype were (HLA-A*01:03, HLA-B*15:10, HLA-B*57:03, HLA-C*03:04, HLA-C*07:01, HLA-C*07:02, and HLA-C*14:02). In HLA-LA, significant alleles were HLA-A*01:03, HLA-A*02:01, HLA-A*03:01, HLA-A*03:05, HLA-B*15:10, HLA-B*27:05, HLA-B*35:01, HLA-B*48:04, HLA-B*78:02, HLA-C*07:01, HLA-C*07:02, HLA-C*07:04, HLA-C*12:02, and HLA-C*14:02).In HLA-Seq2, significant alleles were (HLA-A*01:03, HLA-A*02:34, HLA-A*29:05, HLA-B*15:10, HLA-B*57:01, HLA-C*04:04, HLA-C*07:01, HLA-C*07:02, HLA-C*07:04, and HLA-C*14:02). Most identified alleles had an odds ratio greater than 1, suggested they are associated with increased severity of the condition being studied. An exception is HLA-A*02:01, identified by HLA-LA, which had an odds ratio 0.162, indicating a potential protective effect. Nonetheless, these initial data require correction for p-value calculations to ensure the validity and reliability of the findings.

All data are detailed in Tables (4.2), (4.3), and (3.4) for Hisat-genotype alleles, Tables (4.5), (4.6), and (4.7) for HLA-LA alleles, and Tables (4.8), (4.9), and (4.10) for HLA-Seq2 alleles. Initial significant P-values are highlighted in red for easy identification.

Table 4.2: Hisat-genotype analysis of HLA-A alleles among COVID-19 patients displaying uncorrected p-values.

Allele	Frequency in COVID-19 Patients	Frequency in the Control sample	Odd Ratio	CI 95%	P value
A*01:01	0.222	0.164	1.454	0.685 to 3.050	0.476
A*01:03	0.028	0.000	140.600	11.62 to 1068	<0.0001
A*02:01	0.139	0.150	0.916	0.385 to 2.317	0.960
A*03:01	0.056	0.047	1.190	0.280 to 4.397	0.877
A*11:01	0.028	0.051	0.528	0.0516 to 2.925	0.794
A*23:01	0.028	0.032	0.865	0.084 to 4.794	0.741
A*24:02	0.111	0.101	1.109	0.420 to 2.938	0.935
A*24:03	0.028	0.015	1.841	0.180 to 10.26	0.945
A*26:01	0.056	0.055	1.012	0.239 to 3.739	0.727
A*29:01	0.028	0.024	1.163	0.113 to 6.457	0.692
A*29:02	0.028	0.007	4.232	0.411 to 23.87	0.600
A*30:01	0.028	0.055	0.493	0.0481 to 2.728	0.730
A*30:02	0.056	0.012	4.765	1.120 to 17.79	0.109
A*33:01	0.028	0.033	0.850	0.083 to 4.713	0.757
A*68:01	0.028	0.024	1.157	0.113 to 6.424	0.689
A*34:02	0.028	0.003	10.160	0.979 to 59.11	0.213

Table4.3 : Hisat-genotype analysis of HLA-B alleles among COVID-19 patients displaying uncorrected p-values.

Allele	Frequency in COVID-19 Patients	Frequency in the Control sample	Odd Ratio	CI 95%	P value
B*07:05	0.028	0.011	2.761	0.269 to 15.46	0.829
B*08:01	0.056	0.026	2.195	0.517 to 8.135	0.559
B*13:02	0.028	0.027	1.037	0.183 to 5.299	0.632
B*14:02	0.028	0.032	0.878	0.086 to 4.871	0.727
B*15:03	0.028	0.006	4.992	0.484 to 28.27	0.516
B*15:10	0.028	0.001	58.550	5.315 to 335.8	0.001
B*27:05	0.028	0.004	7.869	0.761 to 45.24	0.308
B*35:01	0.083	0.050	1.735	0.556 to 5.387	0.589
B*35:02	0.056	0.048	1.159	0.273 to 4.284	0.853
B*35:08	0.028	0.061	0.444	0.043 to 2.456	0.636
B*37:01	0.028	0.006	4.887	0.474 to 27.66	0.526
B*38:01	0.056	0.044	1.290	0.304 to 4.768	0.954

B*40:01	0.028	0.022	1.259	0.123 to 6.994	0.735
B*40:06	0.028	0.002	16.710	1.597 to 100.8	0.081
B*41:01	0.028	0.049	0.557	0.054 to 3.082	0.843
B*41:02	0.028	0.022	1.300	0.127 to 7.225	0.752
B*44:03	0.056	0.019	3.101	0.73 to 11.52	0.308
B*49:01	0.083	0.050	1.745	0.56 to 5.420	0.582
B*50:01	0.028	0.053	0.508	0.05 to 2.814	0.758
B*51:01	0.083	0.076	1.107	0.36 to 3.433	0.884
B*52:01	0.028	0.052	0.519	0.051 to 2.872	0.777
B*53:01	0.028	0.021	1.352	0.132 to 7.517	0.774
B*57:01	0.028	0.018	1.595	0.155 to 8.876	0.865
B*57:03	0.083	0.002	42.920	13.42 to 133.9	<0.0001

Table 4.4 : Hisat-genotype analysis of HLA-C alleles among COVID-19 patients displaying uncorrected p-values.

Allele	**Frequency in COVID-19 Patients**	**Frequency in the Control sample**	**Odd ratio**	**CI 95%**	**P value**
C*02:02	0.028	0.041	0.667	0.061 to 3.871	0.947
C*03:04	0.083	0.000	+infinity	7.062 to +infinity	<0.0001
C*04:01	0.139	0.100	1.463	0.579 to 3.870	0.656
C*06:02	0.111	0.083	1.396	0.495 to 4.250	0.787
C*07:01	0.194	0.028	8.569	3.153 to 26.41	<0.0001
C*07:02	0.056	0.000	+infinity	3.817 to +infinity	0.004
C*07:04	0.028	0.000	+infinity	0.901 to +infinity	0.211
C*07:06	0.028	0.000	+infinity	0.901 to +infinity	0.211
C*08:02	0.028	0.000	+infinity	0.901 to +infinity	0.211
C*12:02	0.056	0.007	8.529	1.288 to 55.10	0.088
C*12:03	0.056	0.035	1.659	0.351 to 6.581	0.863
C*14:02	0.056	0.000	+infinity	3.817 to +infinity	0.004
C*15:02	0.028	0.045	0.613	0.056 to 3.479	0.975
C*15:05	0.028	0.000	+infinity	0.901 to +infinity	0.211
C*16:01	0.028	0.003	8.314	0.426 to 158.3	0.525
C*16:04	0.028	0.000	+infinity	0.901 to +infinity	0.211
C*17:03	0.028	0.000	+infinity	0.901 to +infinity	0.211

Table 4.5 : HLA-LA analysis of HLA-A alleles among COVID-19 patients displaying uncorrected p-values.

Allele	Frequency in COVID-19 Patients	Frequency in the Control sample	Odd Ratio	CI 95%	P value
A*01:01	0.139	0.164	0.821	0.345 to 2.078	0.853
A*01:03	0.028	0.000	140.600	11.62 to 1068	<0.0001
A*02:01	0.028	0.150	0.162	0.016 to 0.896	0.040
A*03:01	0.278	0.047	7.780	3.807 to 16.23	<0.0001
A*03:05	0.083	0.000	+infinity	617.2 to +infinity	<0.0001
A*11:01	0.028	0.051	0.528	0.052 to 2.925	0.794
A*23:01	0.028	0.032	0.865	0.084 to 4.794	0.741
A*24:02	0.139	0.101	1.431	0.601 to 3.625	0.638
A*24:03	0.028	0.015	1.841	0.179 to 10.26	0.945
A*26:01	0.028	0.055	0.492	0.048 to 2.722	0.728
A*29:01	0.028	0.024	1.163	0.113 to 6.457	0.692
A*30:01	0.028	0.055	0.493	0.048 to 2.728	0.730
A*30:02	0.056	0.012	4.765	1.120 to 17.79	0.109
A*33:01	0.028	0.033	0.850	0.083 to 4.713	0.757
A*34:02	0.028	0.003	10.160	0.979 to 59.11	0.213
A*68:01	0.028	0.024	1.157	0.113 to 6.424	0.689

Table 4.6 : HLA-LA analysis of HLA-B alleles among COVID-19 patients displaying uncorrected p-values.

Allele	Frequency in COVID-19 Patients	Frequency in the Control sample	Odd Ratio	CI 95%	P value
B*07:02	0.056	0.026	2.195	0.517 to 8.135	0.559
B*07:05	0.028	0.011	2.696	0.262 to 15.09	0.842
B*08:01	0.028	0.026	1.066	0.104 to 5.919	0.646
B*14:02	0.028	0.032	0.878	0.086 to 4.871	0.727
B*15:03	0.028	0.006	4.992	0.484 to 28.27	0.516
B*15:10	0.056	0.001	120.500	25.94 to 511.2	<0.0001
B*27:05	0.111	0.004	34.430	12.80 to 95.93	<0.0001
B*35:01	0.167	0.050	3.817	1.703 to 8.937	0.005
B*35:02	0.028	0.048	0.563	0.055 to 3.118	0.854
B*37:01	0.028	0.006	4.887	0.474 to 27.66	0.526
B*38:01	0.056	0.044	1.290	0.304 to 4.768	0.954
B*40:01	0.028	0.022	1.259	0.123 to 6.994	0.735

B*40:06	0.028	0.002	16.710	1.597 to 100.8	0.081
B*41:02	0.028	0.022	1.300	0.127 to 7.225	0.752
B*44:03	0.028	0.019	1.506	0.147 to 8.379	0.833
B*48:04	0.056	0.000	+infinity	322 to +infinity	< 0.0001
B*51:01	0.083	0.076	1.107	0.355 to 3.433	0.884
B*53:01	0.028	0.021	1.350	0.18 to 9.89	0.774
B*57:01	0.056	0.018	3.280	0.772 to 12.20	0.274
B*57:03	0.028	0.002	13.490	1.81 to 100.3	0.128
B*78:02	0.028	0.000	+infinity	75.93 to +infinity	< 0.0001

Table 4.7 : HLA-LA analysis of HLA-C alleles among COVID-19 patients displaying uncorrected p-values.

Allele	**Frequency in COVID-19 Patients**	**Frequency in the Control sample**	**Odd Ratio**	**CI 95%**	**P value**
C*02:02	0.028	0.041	0.667	0.061 to 3.871	0.947
C*04:01	0.028	0.100	0.259	0.0245 to 1.467	0.272
C*04:04	0.028	0.000	+infinity	0.901 to +infinity	0.211
C*04:29	0.028	0.000	+infinity	0.901 to +infinity	0.211
C*06:02	0.083	0.083	1.015	0.308 to 3.241	0.766
C*06:04	0.028	0.003	8.314	0.426 to 158.3	0.525
C*07:01	0.167	0.028	7.150	2.231 to 23.19	0.001
C*07:06	0.028	0.000	+infinity	0.901 to +infinity	0.211
C*07:02	0.194	0.000	+infinity	16.68 to +infinity	0.000
C*07:04	0.056	0.000	+infinity	3.817 to +infinity	0.004
C*08:02	0.028	0.000	+infinity	0.901 to +infinity	0.211
C*12:02	0.111	0.007	18.130	4.017 to 96.37	0.000
C*12:03	0.056	0.035	1.659	0.351 to 6.581	0.863
C*14:02	0.056	0.000	+infinity	3.817 to +infinity	0.004
C*15:02	0.028	0.045	0.613	0.056 to 3.479	0.974
C*15:05	0.028	0.000	+infinity	0.901 to +infinity	0.211
C*17:01	0.028	0.010	2.752	0.206 to 18.79	0.922

Table 4.8 : Seq2HLA analysis of HLA-A alleles among COVID-19 patients displaying uncorrected p-values.

Allele	Frequency in COVID-19 Patients	Frequency in the Control sample	Odd Ratio	CI 95%	P value
A*01:01	0.167	0.164	1.018	0.455 to 2.377	0.853
A*01:03	0.028	0.000	140.600	11.62 to 1068	<0.0001
A*02:01	0.111	0.150	0.710	0.268 to 1.878	0.677
A*02:34	0.028	0.000	infinity	0.00 to 0.013	<0.0001
A*03:01	0.056	0.047	1.190	0.280 to 4.397	0.877
A*11:01	0.028	0.051	0.528	0.052 to 2.925	0.794
A*24:02	0.056	0.101	0.529	1.231 to 1.925	0.527
A*24:03	0.028	0.015	1.841	0.179 to 10.26	0.945
A*26:01	0.056	0.055	1.012	0.239 to 3.739	0.727
A*29:01	0.028	0.024	1.163	0.113 to 6.457	0.692
A*29:02	0.028	0.007	4.232	0.412 to 23.87	0.600
A*29:05	0.028	0.000	infinity	0.00 to 0.013	<0.0001
A*30:01	0.028	0.055	0.493	0.048 to 2.728	0.730
A*30:02	0.056	0.012	4.765	1.120 to 17.79	0.109
A*32:01	0.083	0.032	2.769	0.9 to 8.444	0.199
A*33:01	0.028	0.033	0.850	0.083 to 4.713	0.757
A*68:01	0.028	0.024	1.157	0.113 to 6.424	0.689
A*34:02	0.028	0.003	10.160	0.979 to 59.11	0.213

Table 4.9: Seq2HLA analysis of HLA-B alleles among COVID-19 patients displaying uncorrected p-values.

Allele	Frequency in COVID-19 Patients	Frequency in the Control sample	Odd Ratio	CI 95%	P value
B*07:05	0.028	0.010	2.761	0.269 to 15.46	0.829
B*08:01	0.056	0.026	2.195	0.517 to 8.135	0.559
B*13:02	0.028	0.027	1.037	0.184 to 5.299	0.632
B*14:02	0.028	0.032	0.878	0.086 to 4.871	0.727
B*15:03	0.028	0.006	4.992	0.484 to 28.27	0.516
B*15:10	0.028	0.000	58.550	5.315 to 335.8	0.001
B*27:05	0.028	0.004	7.869	0.761 to 45.24	0.308
B*35:01	0.083	0.050	1.735	0.556 to 5.387	0.589

B*35:02	0.056	0.048	1.159	0.273 to 4.284	0.853
B*35:08	0.028	0.060	0.444	0.043 to 2.456	0.636
B*37:01	0.028	0.006	4.887	0.474 to 27.66	0.526
B*38:01	0.056	0.044	1.290	0.304 to 4.768	0.954
B*40:01	0.028	0.022	1.259	0.123 to 6.994	0.735
B*40:06	0.028	0.002	16.710	1.597 to 100.8	0.081
B*41:01	0.028	0.049	0.557	0.054 to 3.082	0.843
B*41:02	0.028	0.022	1.300	0.127 to 7.225	0.752
B*44:03	0.056	0.019	3.101	0.73 to 11.52	0.308
B*49:01	0.083	0.050	1.745	0.56 to 5.420	0.582
B*50:01	0.028	0.053	0.508	0.05 to 2.814	0.758
B*51:01	0.083	0.076	1.107	0.355 to 3.433	0.884
B*52:01	0.028	0.052	0.519	0.051 to 2.872	0.777
B*53:01	0.028	0.021	1.352	0.132 to 7.517	0.774
B*57:01	0.028	0.018	1.595	0.155 to 8.876	0.865
B*57:03	0.083	0.002	42.920	13.42 to 133.9	<0.0001

Table 4.10: Seq2HLA analysis of HLA-C alleles among COVID-19 patients displaying uncorrected p-values.

Allele	**Frequency in COVID-19 Patients**	**Frequency in the Control sample**	**Odd ratio**	**CI 95%**	**P value**
C*02:02	0.028	0.003	0.667	0.061 to 3.871	0.947
C*03:03	0.056	0.007	2.804	0.555 to 11.61	0.476
C*03:04	0.028	0.003	+infinity	0.901 to +infinity	0.211
C*04:01	0.083	0.010	0.825	0.252 to 2.552	0.994
C*04:04	0.056	0.007	+infinity	3.817 to +infinity	0.004
C*06:02	0.111	0.014	1.396	0.495 to 4.250	0.787
C*07:01	0.194	0.024	8.569	3.153 to 26.41	<0.0001
C*07:02	0.056	0.007	+infinity	3.817 to +infinity	0.004
C*07:04	0.056	0.007	+infinity	3.817 to +infinity	0.004
C*07:06	0.028	0.003	+infinity	0.901 to +infinity	0.211
C*08:02	0.028	0.003	+infinity	0.901 to +infinity	0.211
C*12:02	0.056	0.007	8.529	1.288 to 55.10	0.088
C*12:03	0.056	0.007	1.659	0.351 to 6.581	0.863
C*14:02	0.056	0.007	+infinity	3.817 to +infinity	0.004
C*15:02	0.028	0.003	0.613	0.056 to 3.479	0.975
C*15:05	0.028	0.003	+infinity	0.901 to +infinity	0.211
C*16:01	0.028	0.003	8.314	0.426 to 158.3	0.525
C*17:03	0.028	0.003	+infinity	0.901 to +infinity	0.211

4.3 The significant HLA alleles among COVID-19 patients

The Benjamini-Hochberg method was used to correct the initial p-values. Each allele was ranked according to its initial p-value within its gene locus. The Benjamini-Hochberg equation was then applied, with the threshold for significance (Pc-value) set at 0.05. The significance levels of these alleles were computed for the three sets derived from the Hisat-genotype, the HLA-LA and the HLA-Seq2 programmes. Listed below in the table (4.11).
All initially suggested significant alleles identified by the different software programs remain significant and statistically associated with COVID-19 severity, as their odds ratios >1 indicated a positive correlation, except HLA-A*02:01, with an odds ratio <1 and a corrected p-value of 0.3224, is considered not significant.

Table 4.11: The significant HLA alleles among COVID-19 patients by Hisat-genotype program with corrected P value.

Hisat-genotype Alleles	P value	Rank	Corrected P-value
A*01:03	< 0.00001	1	0.00016
B*15:10	< 0.0005	2	0.00012
B*57:03	< 0.0001	1	0.00024
C*03:04	<0.0001	1	0.0017
C*07:01	<0.0001	1	0.0017
C*07:02	0.0037	2	0.0314
C*14:02	0.0037	2	0.0314

Table 4.12: The significant HLA alleles among COVID-19 patients by HLA-LA program with corrected P value.

HLA-LA Alleles	P value	Rank	Corrected P value
A*01:03	<0.0001	1	0.0016
A*03:01	<0.0001	1	0.0016
A*03:05	<0.0001	1	0.0016
B*15:10	<0.0001	1	0.0021
B*27:05	<0.0001	1	0.0021

B*35:01	0.0046	2	0.0483
B*48:04	<0.0001	1	0.0021
B*78:02	<0.0001	1	0.0021
C*07:01	0.0005	3	0.0028
C*07:02	0.0001	1	0.0017
C*07:04	0.0037	4	0.0157
C*12:02	0 .0002	2	0.0015
C*14:02	0.0037	4	0.0157

Table 4.13: The significant HLA alleles among COVID-19 patients by Seq2HLA program with corrected P value.

Seq2HLA Alleles	P value	Rank	Corrected P value
A*01:03	<0.0001	1	0.0019
B*15:10	0.0005	2	0.006
B*57:03	<0.0001	1	0.0024
C*07:01	0.0001	1	0.0017
C*07:02	0.004	2	0.034
C*07:04	0.004	2	0.034
C*14:02	0.004	2	0.034

4.4 The T cell epitopes for HLA alleles among Covid-19 patients.

To investigate the potential presence of shared, specifically targeted domains in the SARS-CoV-2 spike and nucleocapsid proteins by the significant HLA alleles identified through the Hisat-genotype software, the Immune Epitope Database (IEDB) was utilized to identify T-cell epitopes associated with these alleles, whether significant or not, as detailed in Appendices A and B. Subsequently, a Python code was developed to explore these epitopes. The identified epitopes were then mapped and visualized for their alignment with the sequences of the SARS-CoV-2 spike and nucleocapsid proteins. This revealed that, among the significant alleles only HLA-C07:02, HLA-C07:01, and HLA-B57:03 had

identified epitopes associated with the SARS-CoV-2 spike protein. Specifically, HLA-C07:02 had seven epitopes, HLA-C07:01 had four, and HLA-B57:03 had one. For the nucleocapsid protein, only HLA-C07:02 and HLA-C07:01 were associated with identified epitopes. HLA-C07:01 exhibited seven epitopes, while HLA-C07:02 had two. as illustrated in figures (4.1) and (4.2).

Notably, there was no common immunogenic domain within the SARS-CoV-2 spike or nucleocapsid proteins shared by the significant alleles or any combination of alleles based on their identified epitopes. However, this does not definitively rule out the existence of such a relationship, as not all epitopes may have been discovered. Utilizing the IEDB, the identified epitopes were located and displayed their relation to the sequence of the spike protein of SARS-COV-2 in the provided links for visualization.

https://drive.google.com/file/d/13KTBNDwKjp4ijngD1IY_MSYfhDRq4cxO/view?usp=sharing

https://drive.google.com/file/d/1OWQ2niEuIn0egbLDJrAk7hVSOGZTDRAP/view?usp=drive_link

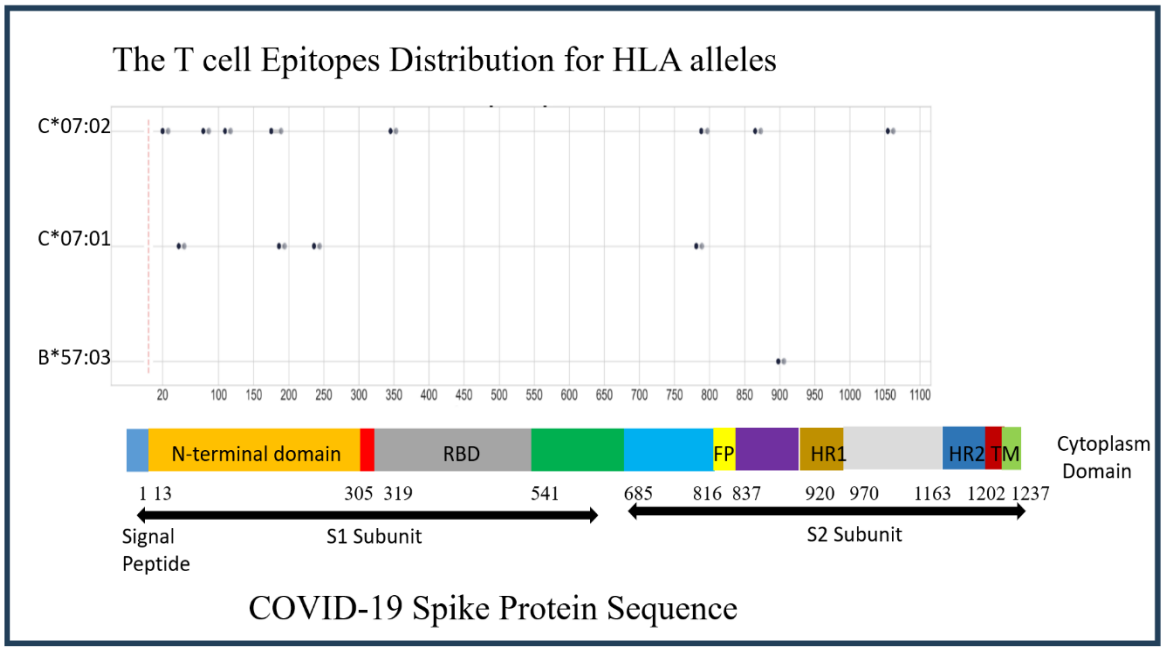

Figure 4.1 : The T cell epitopes for significant HLA alleles in Table (4.11).

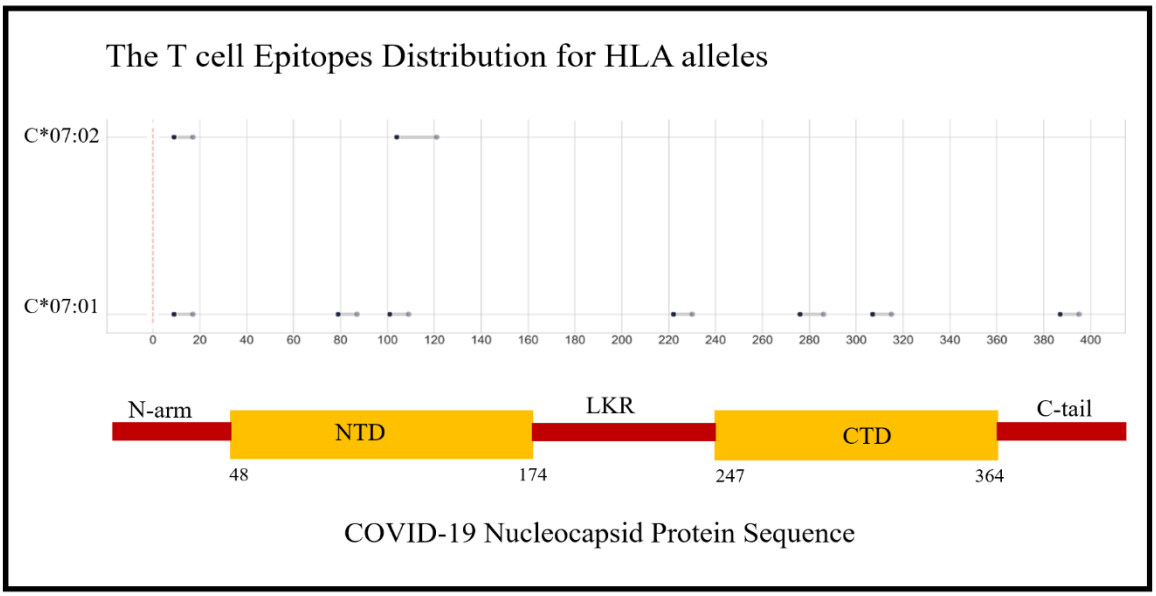

Figure 4.2: The T cell epitopes for significant HLA alleles in Table (4.11).

Moreover, an attempt was made to locate the specific epitopes for each patient's HLA alleles identified in the analysis. However, this effort was limited by the fact that many HLA alleles do not have identified epitopes, preventing a comprehensive alignment picture for individual patients from being captured.

Chapter Five: Discussion and Conclusions

5.1 Statistical Analysis and Results

The genotyping was conducted using three algorithmic software models (HLA-LA, HISAT-genotype, and Seq2HLA), each with varying levels of accuracy. This comparison highlighted differences in the results, as many studies worldwide have not established a unified software standard. HISAT-genotype and Seq2HLA demonstrated consistent significant alleles, including A*01:03, B*15:10, B*57:03, C*03:04, C*07:01, C*07:02, and C*14:02, with Seq2HLA differing only by identifying C*07:04 instead of C*03:04. This consistency is not surprising given their high accuracy rates of 97% and 95%, respectively. In contrast, HLA-LA, which originally boasts 100% accuracy for whole genome sequencing (WGS), showed different significant alleles, such as A*01:03, A*03:01, A*03:05, B*15:10, B*27:05, B*35:01, B*48:04, B*78:02, C*07:01, C*07:02, C*07:04, and C*12:02. However, when HLA-LA was used for whole exome sequencing (WES), its accuracy dropped to 93%, 92%, and 88% for HLA-A, B, and C, respectively, and it showed even lower accuracy than expected compared to the other two programs. This discrepancy could be due to HLA-LA potentially missing non-exonic regulatory or intronic regions crucial for complete HLA variant characterization and performing differently based on the input data type (WES vs. WGS).

Errors surfaced when the same allele in HLA-LA was associated with different variants in other software, as seen with A*03:01 in HLA-LA compared to A*01:01, A*32:01, and A*02:01 in HISAT-genotype for different patients. Despite these differences, the results from HISAT-genotype are considered reliable as they were cross-validated with Seq2HLA, showing only minor discrepancies. This comparison underscores the importance of carefully reviewing literature regarding significant HLA alleles in various studies. When sequencing is performed using NGS, the software's accuracy for variant calling is crucial. The use of three programs demonstrated how accuracy variations could influence the identification of significant HLA alleles. Ultimately, HISAT-genotype was chosen for analysing the second batch of samples due to its high accuracy, reinforcing data robustness.

The relationship between HLA alleles and COVID-19 severity was analyzed using Prism software V5. Odds ratios assessed the strength and direction of associations. where a ratio

less than one indicated a protective factor, suggesting that the allele likely conferred protection against severe COVID-19, as was initially suggested for HLA-A*02:01. However, after correcting the p-value, HLA-A*02:01 was no longer considered significant.

Conversely, an odds ratio exceeding one indicated an increased likelihood of severity, suggesting a positive correlation, with the probability of severity increasing as the odds ratio rose. This was observed for all significant alleles identified in the study. The ultimate assessment of significance relied on correction of p-values that were obtained from Fisher's exact test and Yates' continuity-corrected χ^2 test for categorical data and two-way contingency tables, respectively.

In addition to Odds Ratios, 95% Confidence Intervals were calculated to establish a range wherein the true population parameter likely resides. To address the challenge of multiple comparisons, the Benjamini-Hochberg Equation was applied. This method, known for controlling the false discovery rate, adjusts p-values to mitigate Type I errors in multiple tests. The formula considers both the number of tested alleles P values and their rank, thereby establishing a more stringent measure of significance.

The standard for considering an allele significant was a Corrected P-value below 0.05, a commonly accepted threshold in scientific research for establishing statistical significance. This comprehensive approach, encompassing Odds Ratios, Confidence Intervals, and correction for multiple testing, enhances the reliability and validity of the findings regarding the potential relationship between HLA types and the severity of COVID-19.

An extensive literature review revealed no prior data on the correlation between most HLA alleles and the severity of COVID-19, except for -C*07:01, and HLA-C*14:02. HLA-C*14:02 was highlighted in a Chinese study as significantly associated with severe outcomes in patients (F. Wang et al., 2020b) and is predicted to be a strong binder of SARS-CoV-2 peptides (Barquera et al., 2020). In contrast, HLA-C*07:01 presents contradictory results. An Italian study indicated that HLA-C*07:01 is found in two of the most frequent haplotypes, with one showing a positive association and the other a negative association with COVID-19 severity. (Pisanti et al., 2020).

Additionally, recent research suggests that HLA-C*07:01 is among the genes that may affect vaccine response, linking it to the low-response category. (Santos-Reboucas et al., 2024) In contrast, an in-silico study predicts that HLA-C07:01 is one of the best-fit HLA-I

alleles for binding to SARS-CoV-2 epitopes.(De Moura et al., 2020) However, despite its predicted binding affinity to these epitopes, HLA-C07:01 is still associated with a lower antibody response. B cells do not directly interact with HLA-I molecules for antigen presentation, but cytotoxic T cells activated by HLA-I can influence B cell responses. This suggests that other factors influencing the immune response beyond the HLA system or the specific positioning of the epitopes targeted by HLA-C07:01 may affect the process.

Similarly, for the significant HLA alleles among COVID-19 patients identified through genotyping analysis using the HLA-LA program (Table 4.2), the examination of existing literature has uncovered contradictory information regarding HLA-A*03:01. A study from Moscow involving 111 deceased COVID-19 patients suggested an association with low-risk illness (Shkurnikov et al., 2021). Also, a study in Vietnam indicated that HLA-A*03:01 is more frequent in the control group, suggesting a protective role (Nguyen et al., 2024). Conversely, a study in the UAE found that HLA-A*03:01 had a positive correlation with illness severity (Alnaqbi et al., 2022). Recent research also found that HLA-A*03:01 is significantly associated with serum levels of anti-SARS-CoV-2 antibodies acting as a protection factor (Esposito et al., 2024). while another study on vaccination response indicated a low antibody response (Santos-Reboucas et al., 2024).

A Spanish study proposes that patients with mild disease exhibit Class I HLA molecules with a higher theoretical capacity for binding SARS-CoV-2 peptides and display greater heterozygosity compared to moderate and severe groups. HLA-B*35:01 was frequently found in moderate patients not the severe patients (Iturrieta-zuazo et al., 2020b), while the results of the Palestinians patients suggest a potential role of HLA-B*35:01 that was observed it in two deceased patients as homozygotes (IC30 and IC36), both experiencing severe respiratory failure and requiring ICU admission. with endotracheal tube placement and the other with BIPAP for oxygen supply. Both exhibited D-dimer levels over 7000.

Additionally, it was found in two other ICU patients in heterozygous form for the HLA-B locus, and they also had low oxygen levels. This may suggest a possible association for HLA-B*35:01 in pulmonary dysfunction in COVID-19 patients, supported by the fact that HLA-B*35 contributes to the dysregulated expression of selected ER stress, inflammation, and proliferation-related genes (Lenna et al., 2015). However, this cannot be concluded definitively, as there is contradictory data from the literature. Despite its prediction as a

strong binder to SARS-CoV-2 epitopes (De Moura et al., 2020), HLA-B*35:01's role in disease severity warrants further investigation.

Another allele linked to mild disease is HLA-C*12:02, as indicated by Italian research (Guerini et al., 2022). It is also associated with a good response in producing IgG antibodies against the virus after receiving the Pfizer-BioNTech vaccine (Esposito et al., 2024). For the remaining HLA alleles identified in this research, HLA-C*14:02 is reported to be strongly bound to SARS-CoV-2. There is no further data for other alleles, except the HLA-B*27 which was associated with milder cases (Barquera et al., 2020),

Returning to the HLA-C07:01 allele, as mentioned, An Italian study identified it in two haplotypes: one associated with increased severity and the other with decreased severity (Pisanti et al., 2020). This suggests that HLA-C*07:01 may have a negligible effect on COVID-19 severity. However, the outcomes of any study are heavily influenced by its specific methodology, including the particular inclusion and exclusion criteria for the sample and control groups, as well as whether the comparisons are made to the general population or asymptomatic patients. For instance, the Italian study analyzed regional frequencies of predominant Italian haplotypes from the Italian Bone Marrow Donor Registry. Subsequently, Pearson correlation analyses were executed to examine the correlation between the estimated frequency of regional haplotypes in the population and the incidence and deaths of Covid-19 (Pisanti et al., 2020). It is important to note two points from this study: first, the HLA haplotype distribution in specific regions may be influenced by demographic population differences and their common ancestry; second, the increased infection rate could be attributed to lax quarantine regulations. This highlights how study design can significantly impact the results.

The remaining HLA alleles, with significant p-values, lack supporting evidence in the literature regarding their correlation with COVID-19 severity. Further research is needed to clarify their roles. Additionally, the check-up of B*15:10 reveals a potential source of bias, as it is identified as a homozygote in sample IC01 within the HLA-LA analysis set. This necessitates caution in interpreting the influence of B*15:10 on the study findings. Conversely, an attention is drawn to the notable consistency observed in the association of both C*07:01 and C*07:02 alleles with the severity of COVID-19. These alleles arise as the

most robust indicators, exhibiting a compelling correlation with the manifestation of severe outcomes in the study.

5.2 Epitopes Exploring

Given the absence of supporting evidence in the literature concerning the correlation between COVID-19 severity and the majority of HLA alleles exhibiting a P value less than 0.05, it was important to determine whether these alleles might contribute to an immune response by targeting specific domains within the virus. Considering the Spike protein, a search was conducted in the IEDB database for T cell epitopes susceptible to targeting by the identified HLA alleles to identify any commonalities.

For instance, a crucial cleavage event is required on the Spike protein at a specific site regulated by TMPRSS2 to initiate its membrane fusion activity. If the immune response successfully generates antibodies against this particular site, it could facilitate the activation of membrane fusion by causing cleavage next to the site, thus promoting virus infection. This phenomenon, known as Antibody-Dependent Enhancement (ADE) of virus infection, has been observed in viruses of various families. However, such enhancement has not been reported in SARS, despite being documented in other Coronaviridae family members like Feline Coronavirus (FCoV), which infects both cats and humans. The occurrence of FCoV infection has been associated with the development of a vaccine targeting the S protein.(Huisman et al., 2009)'(Gao et al., 2022)

Regrettably, the investigation did not reveal any commonalities among the epitopes of all the HLA alleles with significant P values neither indicate they are targeting the same virus protein domain. Furthermore, the absence of identified epitopes for some of these specific HLA alleles did not afford a comprehensive understanding of the underlying dynamics.

However, a suggestion indicates that both HLA-C*07:01 and HLA-C*07:02 show the strongest association with the severity of COVID-19 with p value less than one and odd ratio yielded as + infinity. An attention has been directed towards specific epitopes, namely the HLA-C*07:02 epitope (IYKTPPIKDF, spanning amino acids 788 to 797) and the HLA-C*07:01 epitope (VFAQVKQIY, spanning amino acids 781 to 789). Notably, these epitopes exhibit close proximity to the Fusion Peptide (FP) domain (826 – 837) within the spike protein, as represented in figure (4.1).

This FP domain has a critical importance for viral cell entry, being responsible for virion fusion with the target cell membrane. This fusion is facilitated by a cleavage event, in which the internal fusion peptide is exposed, penetrating the target cell membrane and forming an anchor. As it was mentioned earlier in the normal infection process, this cleavage event is triggered by TMPRSS2 at the cleavage site 797R|S798 in SARS-Spike protein S2. Notably, TMPRSS2 must be expressed on the surface of the target cell for viral entry, unless an alternative route is pursued through the endosome-cathepsin L pathway, wherein endocytosis induces S2 cleavage.(Xia, 2021)

A proposition posits that the HLA-C*07:01 and HLA-C*07:02 alleles, as they both target the same region next to position of the Fusion Peptide, may induce a comparable cleavage event. This potential similarity in cleavage events could facilitate viral entry into cells, regardless of TMPRSS2 expression on the cell surface. This hypothesis gains support from the genomic similarity between SARS-CoV and SARS-CoV-2, with previous studies reporting cross-reactivity of antibodies targeting the spike protein S2 in SARS-CoV, causing cytotoxicity in lung epithelial cells.(Lin et al., 2005) However, this suggestion may still as a potential avenue for future research work.

The nucleocapsid (N) protein epitopes have not been as extensively studied as those of the spike protein. Most of the resulting HLA alleles lack identified epitopes for the N protein. However, two significant alleles, HLA-C*07:01 and HLA-C*07:02, have identified epitopes that align with the N protein. These epitopes share commonality residues in two positions. The first shared epitope spans residues 9-17. The second, more intriguing epitope, includes a positively charged arginine at position 107 within the N-terminal domain (NTD), which is critical as it contributes directly to RNA binding. The epitope for HLA-C*07:01 is MKDLSPRWY, spanning residues 101-109, while for HLA-C*07:02, the epitope is LSPRWYFYYLGTGPEAGL, spanning residues 104-121. This region is part of the highly positively charged groove, essential for binding the negatively charged RNA molecule. (Silhan et al., 2020)

This RNA-binding groove is believed to play a pivotal role in forming a higher-order structure, known as a supercoil, which is essential for compactly packaging the viral RNA within the nucleocapsid. The full-length N protein forms dimers, and multiple copies of these

dimers assemble to encapsulate the viral RNA. The supercoil structure is crucial for efficient packaging and stabilization of the RNA.

If these HLA-C molecules successfully present these epitopes to the immune system, they could enhance the process of antibody formation via B cells. The resultant antibodies could theoretically inhibit RNA binding by blocking the interaction between the N protein and viral RNA. This inhibition could prevent proper RNA encapsulation, crucial for viral replication and packaging, thereby impairing nucleocapsid formation and rendering the virus non-infectious. Consequently, this would reduce viral replication and enhance immune clearance.

However, these theoretical considerations also raise the possibility of unintended outcomes, such as a persistent immune response or enhanced dimerization of the N protein via neutralize the positive charge, which could affect virus elimination. These hypotheses remain speculative and require further empirical validation. The current discussion highlights a conceptual framework that underscores the potential immunological strategies targeting the N protein, albeit with several limitations and the need for extensive research to confirm these theories.

5.3 Conclusions

This study employs a statistically robust methodology to identify specific HLA class I alleles that may contribute to the severity of COVID-19 illness. The identified alleles encompass A*01:03, B*15:10, B57:03, C03:04, C*07:01, C*07:02, and C*14:02. Highlights the potential significance of C*07:01 and C*07:02 in the manifestation of severe outcomes. It is noteworthy that the outcomes of HLA analyses are susceptible to variations introduced by the genotyping software analysis programs. Consequently, the observed discrepancies in the relationship between HLA and the disease may be attributed to the influence of these software tools. This underscores the importance of discerning the impact of genotyping software on HLA findings and emphasizes the necessity for cautious interpretation of results in the context of disease severity in COVID-19.

The findings underscore the importance of considering multiple factors, including statistical corrections for multiple comparisons, in HLA association studies. Further investigation and replication studies are warranted to validate these findings and explore the

functional implications of the identified HLA alleles and T-cell epitopes in the context of COVID-19.

5.4 Limitations

Sample Size: The study cohort comprises a relatively modest number of patients (18), raising concerns about the generalizability of the findings to broader populations.

Retrospective Nature: The investigation relies on pre-existing patient records, introducing the possibility of incomplete information or lack of uniformity in data collection.

Scope: The study exclusively focuses on the correlation between HLA types and COVID-19 severity, deliberately omitting consideration of additional potential factors, such as other unknown genes whose relationships with the disease have not yet been identified.

Software Discrepancies: The utilization of various software programs for Next-Generation Sequencing (NGS) analysis introduces the potential for divergent results in HLA genotyping, emphasizing the importance of acknowledging and addressing such discrepancies.

Restricted Applicability: Due to Single-Software Analysis of the second set of samples, which restricts the applicability of the results to broader contexts.

T Cell Epitopes: The exploration of T cell epitopes associated with significant HLA alleles is conducted using the Immune Epitope Database (IEDB). It is essential to recognize that not all pertinent epitopes linked to significant HLA alleles may have been identified, underscoring the necessity for continued research in this domain.

5.5 Recommendation

To systematically investigate the association between human leukocyte antigen (HLA) and COVID-19, it is imperative to assemble a diverse set of patient samples. The call for larger patient cohorts in research work underscores the aspiration for more extensive and representative datasets, aiming to enhance the depth of understanding regarding the relationship between HLA alleles and susceptibility to COVID-19. To effectively establish HLA alleles as significant factors linked to disease susceptibility, the establishment of a standardized procedure for HLA genotyping analysis is crucial. This necessitates a collaborative global effort to ensure consistency and comparability across studies.

Moreover, recognizing the unique genetic makeup of the Palestinian population, there is an urgent need for the development of Palestinian databases dedicated to genetic allele frequencies, with a specific focus on HLA. This initiative is vital for capturing the distinct genetic variations within the Palestinian community, contributing valuable insights to the global understanding of HLA-associated factors in the context of COVID-19 severity.

References

Abdelhafiz, A. S., Ali, A., Fouda, M. A., Sayed, D. M., Kamel, M. M., Kamal, L. M., Khalil, M. A., & Bakry, R. M. (2022). HLA-B*15 predicts survival in Egyptian patients with COVID-19. *Human Immunology*, *83*(1), 10–16. https://doi.org/10.1016/j.humimm.2021.09.007

Abolnezhadian, F., Iranparast, S., Shohan, M., Shokati Eshkiki, Z., Hamed, M., Seyedtabib, M., Nashibi, R., Assarehzadegan, M. A., Mard, S. A., Shayesteh, A. A., Neisi, N., Makvandi, M., Alavi, S. M., & Shariati, G. (2024). Evaluation the frequencies of HLA alleles in moderate and severe COVID-19 patients in Iran: A molecular HLA typing study. *Heliyon*, *10*(7), e28528. https://doi.org/10.1016/j.heliyon.2024.e28528

Abu, A., Naqvi, T., Fatima, K., Mohammad, T., Fatima, U., & Singh, I. K. (2020). Insights into SARS-CoV-2 genome, structure, evolution, pathogenesis and therapies: Structural genomics approach. *BBA - Molecular Basis of Disease*, *165878*. https://doi.org/10.1016/j.bbadis.2020.165878

Alicia, S. (2020). *HLA studies in the context of coronavirus outbreaks*. *April*, 1–5. https://doi.org/10.4414/smw.2020.20248

Alitzel Anzurez 1 2, Izumi Naka 3, Shoji Miki 1, Kaori Nakayama-Hosoya 1, Mariko Isshiki 3, Yusuke Watanabe 3, Midori Nakamura-Hoshi 1, Sayuri Seki 1, Takayuki Matsumura 4, Tomohiro Takano 4, Taishi Onodera 4, Yu Adachi 4, Saya Moriyama 4, Kazutaka Teraha, A. K.-T. (2021). Association of HLA-DRB1*09:01 with severe COVID-19. *HLA Immune Response Genetics*, *98*(1), 37–42. https://doi.org/10.1111/tan.14256

Alnaqbi, H., Tay, G. K., Jelinek, H. F., Francis, A., Alefishat, E., El Haj Chehadeh, S., Tahir Saeed, A., Hussein, M., Laila Salameh, Mahboub, B. H., Uddin, M., Alkaabi, N., & Alsafar, H. S. (2022). HLA repertoire of 115 UAE nationals infected with SARS-CoV-2. *Human Immunology*, *83*(1), 1–9. https://doi.org/10.1016/j.humimm.2021.08.012

Arnaiz-Villena, A., Juarez, I., Suarez-Trujillo, F., López-Nares, A., Vaquero, C., Palacio-Gruber, J., & Martin-Villa, J. M. (2021). HLA-G: Function, polymorphisms and pathology. *International Journal of Immunogenetics*, *48*(2), 172–192. https://doi.org/10.1111/iji.12513

Augusto, D. G., & Hollenbach, J. A. (2022). HLA variation and antigen presentation in COVID-19 and SARS-CoV-2 infection. *Immunology*, *76*(January). https://doi.org/10.1016/j.coi.2022.102178%0A0952-7915/©

Augusto, D. G., Yusufali, T., Sabatino, J. J., Peyser, N. D., Murdolo, D., Butcher, X., Murray, V., Pae, V., Sarvadhavabhatla, S., Gill, G., Lynch, K., Yun, C., Maguire, C., Peluso, M. J., Hoh, R., Henrich, T. J., Deeks, S. G., Davidson, M., Lu, S., … Stephen, R. (2022). *A common allele of HLA mediates asymptomatic SARS-CoV-2 infection*.

Barquera, R., Collen, E., Di, D., Buhler, S., & Sanchez-mazas, A. (2020). *Binding affinities of 438 HLA proteins to complete proteomes of seven pandemic viruses and distributions of strongest and weakest HLA peptide binders in populations worldwide*. *May*, 277–298. https://doi.org/10.1111/tan.13956

Benjam, K., Knoll, R., & Goldman, Jason, et al. (2020). Early IFN-a signatures and persistent dysfunction are distinguishing features of NK cells in severe COVID-19.

Ann Oncol, *January*, 19–21.

Cai Y, Zhang J, Xiao T, Peng H, Sterling SM, Walsh RM, R. S., & Rits-Volloch S, C. B. M. (2020). Distinct conformational states of SARS- CoV-2 spike protein. *Science*, *396*, 1586–1592. http://dx.doi.org/10.1016/j.ndteint.2014.07.001%0Ahttps://doi.org/10.1016/j.ndteint.2017.12.003%0Ahttp://dx.doi.org/10.1016/j.matdes.2017.02.024

Caterina A.M. La Porta, S. Z. (2020). Estimating the Binding of Sars-CoV-2 Peptides to HLA Class I in Human Subpopulations Using Artificial Neural Networks. *Cell Systems*, *11*(January), 1–6. https://doi.org/10.1016/j.cels.2020.08.011

Chen, Z., & John Wherry, E. (2020). T cell responses in patients with COVID-19. *Nature Reviews Immunology*, *20*(9), 529–536. https://doi.org/10.1038/s41577-020-0402-6

Choo, S. Y. (2007). The HLA system: Genetics, immunology, clinical testing, and clinical implications. In *Yonsei Medical Journal* (Vol. 48, Issue 1, pp. 11–23). https://doi.org/10.3349/ymj.2007.48.1.11

Corman, V. M., Albarrak, A. M., Omrani, A. S., Albarrak, M. M., Farah, M. E., Almasri, M., Muth, D., Sieberg, A., Meyer, B., Assiri, A. M., Binger, T., Steinhagen, K., Lattwein, E., Al-Tawfiq, J., Müller, M. A., Drosten, C., & Memish, Z. A. (2015). Viral Shedding and Antibody Response in 37 Patients with Middle East Respiratory Syndrome Coronavirus Infection. *Clinical Infectious Diseases*, *62*(4), 477–483. https://doi.org/10.1093/cid/civ951

De Moura, R. R., Agrelli, A., Santos-Silva, C. A., Silva, N., Assunção, B. R., Brandão, L., Benko-Iseppon, A. M., & Crovella, S. (2020). Immunoinformatic approach to assess

SARS-CoV-2 protein S epitopes recognised by the most frequent MHC-I alleles in the Brazilian population. *Journal of Clinical Pathology*, 1–5. https://doi.org/10.1136/jclinpath-2020-206946

Del Valle, D. M., Kim-Schulze, S., Huang, H.-H., Beckmann, N. D., Nirenberg, S., Wang, B., Lavin, Y., Swartz, T. H., Madduri, D., Stock, A., Marron, T. U., Xie, H., Patel, M., Tuballes, K., Van Oekelen, O., Rahman, A., Kovatch, P., Aberg, J. A., Schadt, E., … Gnjatic, S. (2020). An inflammatory cytokine signature predicts COVID-19 severity and survival. *Nature Medicine*, *3*, 1–8. https://doi.org/10.1038/s41591-020-1051-9

Dilthey, A. T., Mentzer, A. J., Carapito, R., Cutland, C., Cereb, N., Madhi, S. A., Rhie, A., Koren, S., Bahram, S., McVean, G., & Phillippy, A. M. (2019). HLA∗LA - HLA typing from linearly projected graph alignments. *Bioinformatics*, *35*(21), 4394–4396. https://doi.org/10.1093/bioinformatics/btz235

Esposito, M., Minnai, F., Copetti, M., Miscio, G., Perna, R., Piepoli, A., De Vincentis, G., Benvenuto, M., D'Addetta, P., Croci, S., Baldassarri, M., Bruttini, M., Fallerini, C., Brugnoni, R., Cavalcante, P., Baggi, F., Corsini, E. M. G., Ciusani, E., Andreetta, F., … Colombo, F. (2024). Human leukocyte antigen variants associate with BNT162b2 mRNA vaccine response. *Communications Medicine*, *4*(1), 10–16. https://doi.org/10.1038/s43856-024-00490-2

Fan, Y. Y., Huang, Z. T., Li, L., Wu, M. H., Yu, T., Koup, R. A., Bailer, R. T., & Wu, C. Y. (2009). Characterization of SARS-CoV-specific memory T cells from recovered individuals 4 years after infection. *Archives of Virology*, *154*(7), 1093–1099. https://doi.org/10.1007/s00705-009-0409-6

Farahani, R. H., Esmaeilzadeh, E., Asl, A. N., Heidari, M. F., & Hazrati, E. (2021). Frequency of hla alleles in a group of severe covid-19 iranian patients. *Iranian Journal of Public Health*, *50*(9), 1882–1886. https://doi.org/10.18502/ijph.v50i9.7061

Faure, E., Poissy, J., Goffard, A., Fournier, C., Kipnis, E., Titecat, M., Bortolotti, P., Martinez, L., Dubucquoi, S., Dessein, R., Gosset, P., Mathieu, D., & Guery, B. (2014). Distinct immune response in two MERS-CoV-infected patients: Can we go from bench to bedside? *PLoS ONE*, *9*(2). https://doi.org/10.1371/journal.pone.0088716

Finkel, Y., Mizrahi, O., Nachshon, A., Weingarten-Gabbay, S., Morgenstern, D., Yahalom-Ronen, Y., Tamir, H., Achdout, H., Stein, D., Israeli, O., Beth-Din, A., Melamed, S., Weiss, S., Israely, T., Paran, N., Schwartz, M., & Stern-Ginossar, N. (2021). The coding capacity of SARS-CoV-2. *Nature*, *589*(7840), 125–130. https://doi.org/10.1038/s41586-020-2739-1

Gao, Y. yu, Liang, X. yu, Wang, Q., Zhang, S., Zhao, H., Wang, K., Hu, G. xue, Liu, W. J., & Gao, F. shan. (2022). Mind the feline coronavirus: Comparison with SARS-CoV-2. *Gene*, *825*(March), 146443. https://doi.org/10.1016/j.gene.2022.146443

Garcia, K. C., & Adams, E. J. (2005). How the T cell receptor sees antigen - A structural view. *Cell*, *122*(3), 333–336. https://doi.org/10.1016/j.cell.2005.07.015

Georg, P., Astaburuaga-García, R., Bonaguro, L., Brumhard, S., Michalick, L., Lippert, L. J., Kostevc, T., Gäbel, C., Schneider, M., Streitz, M., Demichev, V., Gemünd, I., Barone, M., Tober-Lau, P., Helbig, E. T., Stein, J., Dey, H.-P., Paclik, D., Mülleder, M., … Sawitzki, B. (2021). Complement Activation Induces Excessive T Cell Cytotoxicity in Severe COVID-19. *SSRN Electronic Journal*, *January*.

https://doi.org/10.2139/ssrn.3866835

Gorkhali, R., Koirala, P., Rijal, S., Mainali, A., Baral, A., & Bhattarai, H. K. (2021). Structure and Function of Major SARS-CoV-2 and SARS-CoV Proteins. *Bioinformatics and Biology Insights*, *15*(Figure 1). https://doi.org/10.1177/11779322211025876

Guerini, F. R., Bolognesi, E., Lax, A., Bianchi, L. N. C., Caronni, A., Zanzottera, M., Agliardi, C., Albergoni, M. P., Banfi, P. I., Navarro, J., & Clerici, M. (2022). HLA Allele Frequencies and Association with Severity of COVID-19 Infection in Northern Italian Patients. *Cells*, *11*(11), 1–8. https://doi.org/10.3390/cells11111792

Gupta, A., Madhavan, M. V., Sehgal, K., Nair, N., Mahajan, S., Sehrawat, T. S., Bikdeli, B., Ahluwalia, N., Ausiello, J. C., Wan, E. Y., Freedberg, D. E., Kirtane, A. J., Parikh, S. A., Maurer, M. S., Nordvig, A. S., Accili, D., Bathon, J. M., Mohan, S., Bauer, K. A., … Landry, D. W. (2020). Extrapulmonary manifestations of COVID-19. *Nature Medicine*, *26*(7), 1017–1032. https://doi.org/10.1038/s41591-020-0968-3

Hardenbrook, N. J., & Zhang, P. (2022). A structural view of the SARS-CoV-2 virus and its assembly. *Current Opinion in Virology*, *52*, 123–134. https://doi.org/10.1016/j.coviro.2021.11.011

Hartenian, E., Nandakumar, D., Lari, A., Ly, M., Tucker, J. M., & Glaunsinger, B. A. (2020). The molecular virology of coronaviruses. *Journal of Biological Chemistry*, *295*(37), 12910–12934. https://doi.org/10.1074/jbc.REV120.013930

Hoffmann, M., Hofmann-Winkler, H., Smith, J. C., Krüger, N., Arora, P., Sørensen, L. K., Søgaard, O. S., Hasselstrøm, J. B., Winkler, M., Hempel, T., Raich, L., Olsson, S.,

Danov, O., Jonigk, D., Yamazoe, T., Yamatsuta, K., Mizuno, H., Ludwig, S., Noé, F., … Pöhlmann, S. (2021). Camostat mesylate inhibits SARS-CoV-2 activation by TMPRSS2-related proteases and its metabolite GBPA exerts antiviral activity. *EBioMedicine*, *65*(Ivm). https://doi.org/10.1016/j.ebiom.2021.103255

Hoffmann, M., Kleine-Weber, H., Schroeder, S., Krüger, N., Herrler, T., Erichsen, S., Schiergens, T. S., Herrler, G., Wu, N. H., Nitsche, A., Müller, M. A., Drosten, C., & Pöhlmann, S. (2020). SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell*, *181*(2), 271-280.e8. https://doi.org/10.1016/j.cell.2020.02.052

Howard, F. H. N., Kwan, A., Winder, N., Mughal, A., Collado-Rojas, C., & Muthana, M. (2022). Understanding Immune Responses to Viruses—Do Underlying Th1/Th2 Cell Biases Predict Outcome? *Viruses*, *14*(7). https://doi.org/10.3390/v14071493

Huisman, W., Martina, B. E. E., Rimmelzwaan, G. F., Gruters, R. A., & Osterhaus, A. D. M. E. (2009). Vaccine-induced enhancement of viral infections. *Vaccine*, *27*(4), 505–512. https://doi.org/10.1016/j.vaccine.2008.10.087

Iturrieta-zuazo, I., Geraldine, C., García-soidán, A., Pintos-fonseca, A. D. M., Alonso-alarcón, N., & Pariente-rodríguez, R. (2020a). Possible role of HLA class-I genotype in SARS-CoV-2 infection and progression: A pilot study in a cohort of Covid-19 Spanish patients. *Clinical Immunology Journal*, *219*(January).

Iturrieta-zuazo, I., Geraldine, C., García-soidán, A., Pintos-fonseca, A. D. M., Alonso-alarcón, N., & Pariente-rodríguez, R. (2020b). Possible role of HLA class-I genotype in SARS-CoV-2 infection and progression: A pilot study in a cohort of Covid-19

Spanish patients. *Clincal Immunology*, *219*(January).

Kai, H., & Kai, M. (2020). Interactions of coronaviruses with ACE2, angiotensin II, and RAS inhibitors—lessons from available evidence and insights into COVID-19. *Hypertension Research*, *43*(7), 648–654. https://doi.org/10.1038/s41440-020-0455-8

Khailany, R. A., Safdar, M., & Ozaslan, M. (2020). Genomic characterization of a novel SARS-CoV-2 Rozhgar. *Gene Reports*, *19/ 100682*. https://doi.org/10.1016/j.genrep.2020.100682

Kudlay, D., Kofiadi, I., & Khaitov, M. (2022). Peculiarities of the T Cell Immune Response in COVID-19. *Vaccines*, *10*(2), 1–16. https://doi.org/10.3390/vaccines10020242

Lan, J., Ge, J., Yu, J., Shan, S., Zhou, H., Fan, S., & Zhang, Q. (2020). Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature*, *581*(May). https://doi.org/10.1038/s41586-020-2180-5

Lenna, S., Assassi, S., Farina, G. A., Mantero, J. C., Scorza, R., Lafyatis, R., Farber, H. W., & Trojanowska, M. (2015). The HLA-B*35 allele modulates ER stress, inflammation and proliferation in PBMCs from Limited Cutaneous Systemic Sclerosis patients. *Arthritis Research and Therapy*, *17*(1), 1–10. https://doi.org/10.1186/s13075-015-0881-1

Li, Q., Wang, Y., Sun, Q., Knopf, J., Herrmann, M., Lin, L., Jiang, J., Shao, C., Li, P., He, X., Hua, F., Niu, Z., Ma, C., Zhu, Y., Ippolito, G., Piacentini, M., Estaquier, J., Melino, S., Weiss, F. D., … Shi, Y. (2022). Immune response in COVID-19: what is next? *Cell Death and Differentiation*, *29*(6), 1107–1122. https://doi.org/10.1038/s41418-022-01015-x

Lin, Y. S., Lin, C. F., Fang, Y. T., Kuo, Y. M., Liao, P. C., Yeh, T. M., Hwa, K. Y., Shieh, C. C. K., Yen, J. H., Wang, H. J., Su, I. J., & Lei, H. Y. (2005). Antibody to severe acute respiratory syndrome (SARS)-associated coronavirus spike protein domain 2 cross-reacts with lung epithelial cells and causes cytotoxicity. *Clinical and Experimental Immunology*, *141*(3), 500–508. https://doi.org/10.1111/j.1365-2249.2005.02864.x

Liu, P., Yao, M., Gong, Y., Song, Y., Chen, Y., Ye, Y., Liu, X., Li, F., Dong, H., Meng, R., Chen, H., & Zheng, A. (2021). Benchmarking the Human Leukocyte Antigen Typing Performance of Three Assays and Seven Next-Generation Sequencing-Based Algorithms. *Frontiers in Immunology*, *12*(March), 1–11. https://doi.org/10.3389/fimmu.2021.652258

Malik, Y. A. (2020). Properties of Coronavirus and SARS-CoV-2. *Malays J Pathol*, *42*(1), 3–11.

Malone, B., Urakova, N., Snijder, E. J., & Campbell, E. A. (2022). Structures and functions of coronavirus replication–transcription complexes and their relevance for SARS-CoV-2 drug design. *Nature Reviews Molecular Cell Biology*, *23*(1), 21–39. https://doi.org/10.1038/s41580-021-00432-z

Manor, S., Halagan, M., Shriki, N., Yaniv, I., Zisser, B., Maiers, M., Madbouly, A., & Stein, J. (2016). High-resolution HLA A~B~DRB1 haplotype frequencies from the Ezer Mizion Bone Marrow Donor Registry in Israel. *Human Immunology*, *77*(12), 1114–1119. https://doi.org/10.1016/j.humimm.2016.09.004

Marchal, A., Cirulli, E. T., Neveux, I., Bellos, E., Thwaites, R. S., Barrett, K. M. S., Zhang,

Y., Nemes-bokun, I., Kalinova, M., Tangye, S. G., Spaan, A. N., Lack, J. B., Ghosn, J., Gorochov, G., Tubach, F., Hausfater, P., & Genetic, C. H. (2024). Lack of association between classical HLA genes and asymptomatic SARS-CoV-2 infection. *Human Genetics and Genomics Advances*, 100300. https://doi.org/10.1016/j.xhgg.2024.100300

Mariuzza, R. A., Agnihotri, P., & Orban, J. (2020). The structural basis of T-cell receptor (TCR) activation: An enduring enigma. *Journal of Biological Chemistry*, *295*(4), 914–925. https://doi.org/10.1074/jbc.REV119.009411

Marongiu, L., Valache, M., Facchini, F. A., & Granucci, F. (2021). How dendritic cells sense and respond to viral infections. *Clinical Science*, *135*(19), 2217–2242. https://doi.org/10.1042/CS20210577

Medhasi, S., & Chantratita, N. (2022). Human Leukocyte Antigen (HLA) System: Genetics and Association with Bacterial and Viral Infections. *Journal of Immunology Research*, *2022*. https://doi.org/10.1155/2022/9710376

Min, C. K., Cheon, S., Ha, N. Y., Sohn, K. M., Kim, Y., Aigerim, A., Shin, H. M., Choi, J. Y., Inn, K. S., Kim, J. H., Moon, J. Y., Choi, M. S., Cho, N. H., & Kim, Y. S. (2016). Comparative and kinetic analysis of viral shedding and immunological responses in MERS patients representing a broad spectrum of disease severity. *Scientific Reports*, *6*(May), 1–12. https://doi.org/10.1038/srep25359

Mollica, V., Rizzo, A., & Massari, F. (2020). The pivotal role of TMPRSS2 in coronavirus disease 2019 and prostate cancer. *Future Oncology*, *16*(27), 2029–2033. https://doi.org/10.2217/fon-2020-0571

Molodtsov, I. A., Kegeles, E., Mitin, A. N., Mityaeva, O., Musatova, O. E., Panova, A. E., Pashenkov, M. V, Peshkova, I. O., Almaqdad, A., Asaad, W., Budikhina, A. S., Deryabin, A. S., Dolzhikova, I. V, Filimonova, I. N., Gracheva, A. N., Ivanova, O. I., Kizilova, A., Komogorova, V. V, Komova, A., … Vasilieva, E. (2021). A prospective study of the protective effect of SARS-CoV-2–specific antibodies and T cells in Moscow residents. *MedRxiv*, *PG-2021.08.19.21262278*, 2021.08.19.21262278. http://medrxiv.org/content/early/2021/08/22/2021.08.19.21262278.abstract NS -

Montazersaheb, S., Hosseiniyan Khatibi, S. M., Hejazi, M. S., Tarhriz, V., Farjami, A., Ghasemian Sorbeni, F., Farahzadi, R., & Ghasemnejad, T. (2022). COVID-19 infection: an overview on cytokine storm and related interventions. *Virology Journal*, *19*(1), 1–15. https://doi.org/10.1186/s12985-022-01814-1

MOORE, P. E. (1955). *World Health Organization.* Canadian Services Medical Journal. https://doi.org/10.4135/9781452218564.n760

Naserghandi, A., Allameh, S. F., & Saffarpour, R. (2020). All about COVID-19 in brief. *New Microbes and New Infections*, *35*, 100678. https://doi.org/10.1016/j.nmni.2020.100678

Ng, O. W., Chia, A., Tan, A. T., Jadi, R. S., Leong, H. N., Bertoletti, A., & Tan, Y. J. (2016). Memory T cell responses targeting the SARS coronavirus persist up to 11 years post-infection. *Vaccine*, *34*(17), 2008–2014. https://doi.org/10.1016/j.vaccine.2016.02.063

Nguyen, A., Julianne K. David, S. K. M., Nellore, A., Reid F. Thompsona, Mary A. Wood, B. R. W., & AComputational. (2020). Human Leukocyte Antigen Susceptibility Map

for Severe Acute Respiratory Syndrome Coronavirus 2. *Journal of Virology*, *94*(13), 1–12.

Nguyen et al. (2024). HLA alleles associated with susceptibility and severity of the COVID-19 in Vietnamese. *Human Immunology*. https://doi.org/10.1016/j.humimm.2024.110796

Nishiga, M., Wang, D. W., Han, Y., Lewis, D. B., & Wu, J. C. (2020). COVID-19 and cardiovascular disease: from basic mechanisms to clinical perspectives. *Nature Reviews Cardiology*, *17*(9), 543–558. https://doi.org/10.1038/s41569-020-0413-9

Novelli, A., Andreani, M., Biancolella, M., Liberatoscioli, L., Passarelli, C., Colona, V. L., Rogliani, P., Leonardis, F., Campana, A., Carsetti, R., Andreoni, M., Bernardini, S., Novelli, G., & Locatelli, F. (2020). HLA allele frequencies and susceptibility to COVID-19 in a group of 99 Italian patients. *Hla*, *96*(5), 610–614. https://doi.org/10.1111/tan.14047

Palestinian MOH. (n.d.). *Covid 19 in Palestine*. Palestinain MOH Covid 19 Reports. https://corona.ps/

Palestinian National Scientific Committe. (n.d.). Palestinain National Covid-19 Management Protocol. *Palestinain MOH Website*. https://site.moh.ps/index/Books/BookType/2/Language/ar

Pisanti, S., Deelen, J., Gallina, A. M., Caputo, M., Citro, M., Abate, M., Sacchi, N., Vecchione, C., & Martinelli, R. (2020). Correlation of the two most frequent HLA haplotypes in the Italian population to the differential regional incidence of Covid-19. *Journal of Translational Medicine*, *18*(1), 1–16. https://doi.org/10.1186/s12967-020-

02515-5

Qutob, N., Salah, Z., Richard, D., Darwish, H., Sallam, H., Shtayeh, I., Najjar, O., Ruzayqat, M., Najjar, D., Balloux, F., & van Dorp, L. (2021). Genomic epidemiology of the first epidemic wave of severe acute respiratory syndrome coronavirus 2 (Sars-cov-2) in Palestine. *Microbial Genomics*, *7*(6). https://doi.org/10.1099/MGEN.0.000584

Renhong Yan, Yuanyuan Zhang1, Yaning Li, Lu Xia, Yingying Guo, Q. Z. (2020). Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science*, *367*(March), 1444–1448.

Rydyznski Moderbacher, C., Ramirez, S. I., Dan, J. M., Grifoni, A., Hastie, K. M., Weiskopf, D., Belanger, S., Abbott, R. K., Kim, C., Choi, J., Kato, Y., Crotty, E. G., Kim, C., Rawlings, S. A., Mateus, J., Tse, L. P. V., Frazier, A., Baric, R., Peters, B., … Crotty, S. (2020). Antigen-Specific Adaptive Immunity to SARS-CoV-2 in Acute COVID-19 and Associations with Age and Disease Severity. *Cell*, *183*(4), 996-1012.e19. https://doi.org/10.1016/j.cell.2020.09.038

Salvi, V., Nguyen, H. O., Sozio, F., Schioppa, T., Gaudenzi, C., Laffranchi, M., Scapini, P., Passari, M., Barbazza, I., Tiberio, L., Tamassia, N., Garlanda, C., Del Prete, A., Cassatella, M. A., Mantovani, A., Sozzani, S., & Bosisio, D. (2021). SARS-CoV-2–associated ssRNAs activate inflammation and immunity via TLR7/8. *JCI Insight*, *6*(18), 1–15. https://doi.org/10.1172/jci.insight.150542

Sánchez-Velasco, P., Karadsheh, N. S., García-Martín, A., Ruíz De Alegría, C., & Leyva-Cobián, F. (2001). Molecular analysis of HLA allelic frequencies and haplotypes in

Jordanians and comparison with other related populations. *Human Immunology*, *62*(9), 901–909. https://doi.org/10.1016/S0198-8859(01)00289-0

Santos-Reboucas, C. B., Ferreira, C. D. S., Nogueira, J. de S., Brustolini, O. J., de Almeida, L. G. P., Gerber, A. L., Guimaraes, A. P. de C., Piergiorge, R. M., Struchiner, C. J., Porto, L. C., & de Vasconcelos, A. T. R. (2024). Immune response stability to the SARS-CoV-2 mRNA vaccine booster is influenced by differential splicing of HLA genes. *Sci Rep*, 1–14. https://doi.org/10.1038/s41598-024-59259-1

Schulte-Schrepping, J., Reusch, N., Paclik, D., Baßler, K., Schlickeiser, S., Zhang, B., Krämer, B., Krammer, T., Brumhard, S., Bonaguro, L., De Domenico, E., Wendisch, D., Grasshoff, M., Kapellos, T. S., Beckstette, M., Pecht, T., Saglam, A., Dietrich, O., Mei, H. E., … Ziebuhr, J. (2020). Severe COVID-19 Is Marked by a Dysregulated Myeloid Cell Compartment. *Cell*, *182*(6), 1419-1440.e23. https://doi.org/10.1016/j.cell.2020.08.001

Seliger, B., Ritz, U., & Ferrone, S. (2006). Molecular mechanisms of HLA class I antigen abnormalities following viral infection and transformation. *International Journal of Cancer*, *118*(1), 129–138. https://doi.org/10.1002/ijc.21312

Shafqat, A., Shafqat, S., Salameh, S. Al, Kashir, J., Alkattan, K., & Yaqinuddin, A. (2022). Mechanistic Insights Into the Immune Pathophysiology of COVID-19; An In-Depth Review. *Frontiers in Immunology*, *13*(March), 1–25. https://doi.org/10.3389/fimmu.2022.835104

Shkurnikov, M., Nersisyan, S., Jankevic, T., Galatenko, A., Gordeev, I., Vechorko, V., & Tonevitsky, A. (2021). Association of HLA Class I Genotypes With Severity of

Coronavirus Disease-19. *Frontiers in Immunology*, *12*(February). https://doi.org/10.3389/fimmu.2021.641900

Silhol, F., Sarlon, G., Deharo, J. C., & Vaïsse, B. (2020). Downregulation of ACE2 induces overstimulation of the renin–angiotensin system in COVID-19: should we block the renin–angiotensin system? *Hypertension Research*, *43*(8), 854–856. https://doi.org/10.1038/s41440-020-0476-3

Spinetti, T., Hirzel, C., Fux, M., Walti, L. N., Schober, P., Stueber, F., Luedi, M. M., & Schefold, J. C. (2020). Reduced monocytic HLA-DR expression indicates immunosuppression in critically ill COVID-19 patients. *Anesthesia & Analgesia*, *Publish Ah*(4), 993–999. https://doi.org/10.1213/ane.0000000000005044

State of Palestine. (2020). State of Emergency: Palestine's COVID-19 Response Plan. *World Health Organisation*, 8. http://www.emro.who.int/images/stories/palestine/documents/Palestine_Authority_COVID-19_Response_Plan_Final_26_3_2020.pdf?ua=1

Szeto, C., Lobos, C. A., Nguyen, A. T., & Gras, S. (2021). *TCR Recognition of Peptide–MHC-I Rule Makers and Breakers*.

Tan, D., Kang, N., Zhu, Y., Hou, J., Wang, H., Xu, H., Zu, C., Gao, Z., Liu, M., Liu, N., Deng, Q., Lu, H., Liu, J., & Xie, Y. (2024). Construction and efficacy testing of DNA vaccines containing HLA-A*02:01-restricted SARS-CoV-2 T-cell epitopes predicted by immunoinformatics. *Acta Biochimica et Biophysica Sinica*, *56*(April), 1–11. https://doi.org/10.3724/abbs.2024039

Tay, G. K., Alnaqbi, H., Chehadeh, S., Peramo, B., Mustafa, F., Rizvi, T. A., Mahboub, B.

H., Uddin, M., Alkaabi, N., Alefishat, E., Jelinek, H. F., & Alsafar, H. (2023). HLA class I associations with the severity of COVID-19 disease in the United Arab Emirates. *PLoS ONE*, *18*(9 September), 1–14. https://doi.org/10.1371/journal.pone.0285712

Thuesen, N. H., Klausen, M. S., Gopalakrishnan, S., Trolle, T., & Renaud, G. (2022). Benchmarking freely available HLA typing algorithms across varying genes, coverages and typing resolutions. *Frontiers in Immunology*, *13*(November), 1–15. https://doi.org/10.3389/fimmu.2022.987655

University, J. H. (2020). *Center for Systems Science and Engineering at Johns Hopkins University*. COVID-19 Dashboard. https://coronavirus.jhu.edu/map.html

Vigón, L., Galán, M., Torres, M., Martín-Galiano, A. J., Rodríguez-Mora, S., Mateos, E., Corona, M., Malo, R., Navarro, C., Murciano-Antón, M. A., García-Gutiérrez, V., Planelles, V., Martínez-Laso, J., López-Huertas, M. R., Coiras, M., Herrador, E. A., Alemany, P. A., Martos, V. B., Chamorro, S., … Rayado, A. V. (2022). Association between HLA-C alleles and COVID-19 severity in a pilot study with a Spanish Mediterranean Caucasian cohort. *PLoS ONE*, *17*(8 August), 1–20. https://doi.org/10.1371/journal.pone.0272867

Wajnberg, A., Amanat, F., Firpo, A., Altman, D. R., Bailey, M. J., Mansour, M., McMahon, M., Meade, P., Mendu, D. R., Muellers, K., Stadlbauer, D., Stone, K., Strohmeier, S., Simon, V., Aberg, J., Reich, D. L., Krammer, F., & Cordon-Cardo, C. (2020). Robust neutralizing antibodies to SARS-CoV-2 infection persist for months. *Science*, *370*(6521), 1227–1230. https://doi.org/10.1126/science.abd7728

Wang, F., Huang, S., Gao, H., Zhou, Y., Lai, C., Li, Z., Xian, W., Qian, X., Li, Z., Huang, Y., Tang, Q., Liu, P., Chen, R., Liu, R., Li, X., Xin, T., Xuan, Z., Bai, Y., Duan, G., … Liu, L. (2020a). Initial Study of Human Genetic Contribution to COVID-19 Severity and Susceptibility. *MedRxiv*, 2020.06.09.20126607. http://medrxiv.org/content/early/2020/06/11/2020.06.09.20126607.abstract

Wang, F., Huang, S., Gao, R., Zhou, Y., Lai, C., Li, Z., Xian, W., Qian, X., Li, Z., Huang, Y., Tang, Q., Liu, P., Chen, R., Liu, R., Li, X., Tong, X., Zhou, X., Bai, Y., Duan, G., … Liu, L. (2020b). Initial whole-genome sequencing and analysis of the host genetic contribution to COVID-19 severity and susceptibility. *Cell Discovery*, *6*(1). https://doi.org/10.1038/s41421-020-00231-4

Wang, M. Y., Zhao, R., Gao, L. J., Gao, X. F., Wang, D. P., & Cao, J. M. (2020). SARS-CoV-2: Structure, Biology, and Structure-Based Therapeutics Development. *Frontiers in Cellular and Infection Microbiology*, *10*(November), 1–17. https://doi.org/10.3389/fcimb.2020.587269

Wang Q, Zhang Y, Wu L, Niu S, Song C, Zhang Z, Lu G, Qiao C, Hu Y, Yuen KY, Wang Q, Zhou H, Yan J, Q. J. (2020). Structural and Functional Basis of SARS-CoV-2 Entry by Using Human ACE2. *Cell*, *181*, 894–904. https://doi.org/10.1016/j.cell.2020.03.045

Wang, W., Zhang, W., Zhang, J., He, J., & Zhu, F. (2020). Distribution of HLA allele frequencies in 82 Chinese individuals with coronavirus disease-2019 (COVID-19). *Hla*, *96*(2), 194–196. https://doi.org/10.1111/tan.13941

Weiner, J., Suwalski, P., Holtgrewe, M., Rakitko, A., Thibeault, C., Müller, M., Patriki, D.,

Quedenau, C., Krüger, U., Ilinsky, V., Popov, I., Balnis, J., Jaitovich, A., Helbig, E. T., Lippert, L. J., Stubbemann, P., Real, L. M., Macías, J., Pineda, J. A., … Heidecker, B. (2021). Increased risk of severe clinical course of COVID-19 in carriers of HLA-C*04:01. *EClinicalMedicine*, *40*. https://doi.org/10.1016/j.eclinm.2021.101099

Wenting Tan, Yanqiu Lu, Juan Zhang, Jing Wang, Yunjie Dan, Zhaoxia Tan, Xiaoqing He, Chunfang Qian, Qiangzhong Sun, Qingli Hu, Honglan Liu, Sikuan Ye, Xiaomei Xiang, Yi Zhou, Wei Zhang, Yanzhi Guo, Xiu-Hua Wang, Weiwei He, Xing Wan, … Guohong Deng. (2020). Viral Kinetics and Antibody Responses in Patients with COVID-19. *MedRxiv*.

WHO. (2020). *Health topics/Coronavirus*. World Health Organization. https://www.who.int/health-topics/coronavirus#tab=tab_3

World Health Organization. (n.d.). *Coronavirus*. https://www.who.int/emergencies/diseases/novel-coronavirus-2019

Xia, X. (2021). Domains and functions of spike protein in sars-cov-2 in the context of vaccine design. *Viruses*, *13*(1), 1–16. https://doi.org/10.3390/v13010109

Xing, Yu-Han; Ni, Wei; Wu, Qin; Li, Wen-Jie; Li, Guo-Ju; Wang, Wen-Di; Tong, Jian-Ning; Song, Xiu-Feng; Wong, Gary Wing-Kin; Xing, Q.-S. (2020). SARS-CoV-2 exacerbates proinflammatory responses in myeloid cells through C-type lectin receptors and Tweety family member 2. *Ann Oncol*, *7*(May), 19–21.

Yang, H., & Rao, Z. (2021). Structural biology of SARS-CoV-2 and implications for therapeutic development. *Nature Reviews Microbiology*, *19*(11), 685–700. https://doi.org/10.1038/s41579-021-00630-8

Yang, L. T., Peng, H., Zhu, Z. L., Li, G., Huang, Z. T., Zhao, Z. X., Koup, R. A., Bailer, R. T., & Wu, C. Y. (2006). Long-lived effector/central memory T-cell responses to severe acute respiratory syndrome coronavirus (SARS-CoV) S antigen in recovered SARS patients. *Clinical Immunology*, *120*(2), 171–178. https://doi.org/10.1016/j.clim.2006.05.002

Ye, Q., Wang, B., & Mao, J. (2020). The pathogenesis and treatment of the 'Cytokine Storm' in COVID-19. *Journal of Infection Journal*, *80*(April), 607–613. https://doi.org/10.1016/j.jinf.2020.03.037

Zhang, F., Gan, R., Zhen, Z., Hu, X., Li, X., Zhou, F., Liu, Y., Chen, C., Xie, S., Zhang, B., Wu, X., & Huang, Z. (2020). Adaptive immune responses to SARS-CoV-2 infection in severe versus mild individuals. *Signal Transduction and Targeted Therapy*, *5*(1). https://doi.org/10.1038/s41392-020-00263-y

Zhang, Y., Xu, J., Jia, R., Yi, C., Gu, W., Liu, P., Dong, X., Zhou, H., Shang, B., Cheng, S., Sun, X., Ye, J., Li, X., Zhang, J., Ling, Z., Ma, L., Wu, B., Zeng, M., Zhou, W., & Sun, B. (2020). Protective humoral immunity in SARS-CoV-2 infected pediatric patients. *Cellular and Molecular Immunology*, *17*(7), 768–770. https://doi.org/10.1038/s41423-020-0438-3

Zhao, J. et al. (2017). Recovery from the Middle East Respiratory Syndrome is Associated with Antibody and T Cell Responses. *Science Immunology*, *176*(10), 139–148. https://doi.org/10.1126/sciimmunol.aan5393.Recovery

Zhu, X., Mannar, D., Srivastava, S. S., Berezuk, A. M., Demers, J. P., Saville, J. W., Leopold, K., Li, W., Dimitrov, D. S., Tuttle, K. S., Zhou, S., Chittori, S., &

Subramaniam, S. (2021). Cryo-electron microscopy structures of the N501Y SARS-CoV-2 spike protein in complex with ACE2 and 2 potent neutralizing antibodies. *PLoS Biology*, *19*(4), 1–17. https://doi.org/10.1371/journal.pbio.3001237

Zipeto, D., Palmeira, J. da F., Argañaraz, G. A., & Argañaraz, E. R. (2020). ACE2/ADAM17/TMPRSS2 Interplay May Be the Main Risk Factor for COVID-19. *Frontiers in Immunology*, *11*(October), 1–10. https://doi.org/10.3389/fimmu.2020.576745

Ziv, O., Price, J., Shalamova, L., Kamenova, T., Goodfellow, I., Weber, F., & Miska, E. A. (2020). The Short- and Long-Range RNA-RNA Interactome of SARS-CoV-2. *Molecular Cell*, *80*(6), 1067-1077.e5. https://doi.org/10.1016/j.molcel.2020.11.004

Zmijewski, J. W., & Pittet, J.-F. (2020). HLA-DR Deficiency and Immunosuppression-Related End-Organ Failure in SARS-CoV2 Infection. *Anesthesia & Analgesia*, *Publish Ah*(Xxx), 1–4. https://doi.org/10.1213/ane.0000000000005140

Zuo, Y., Zuo, M., Yalavarthi, S., Gockman, K., Madison, J. A., Shi, H., Woodard, W., Lezak, S. P., Lugogo, N. L., Knight, J. S., & Kanthi, Y. (2021). Neutrophil extracellular traps and thrombosis in COVID-19. *Journal of Thrombosis and Thrombolysis*, *51*(2), 446–453. https://doi.org/10.1007/s11239-020-02324-z

Appendices

Appendix A: The Study Sample Distribution and patient's Clinical test results.

pt. code	Gender	Age	Hospital	Date of illness	SpO2	O2 therapy	ICU admission	fever	Blood pressure	pulse (60 -100)
IC20	M	37	Hugochavies	10\3\2021	85%	15 L N.C -15L NRM	yes	38.5	N TO 133\79	Normal
IC22	F	47	Hugochavies	3/21/2021	86%	15 L NRFM	Yes	NO	165\95	Rapid
IC23	F	49	Al hilal	30\3\2021	79%	15 L NRFM	Yes, respiratory failure	NO	122\84	Normal
IC24	F	55	Al hilal	2\4\2021	82%	15 L High flow	Yes, respiratory failure	NO		
IC49*	F	49	Hugochavies	21\3\2021	83%	15L NRFM	YES	38	94\50	
IC26*	M	27	Alaskary	30\3\2021	65%	10 L FM – 15 L NRFM - CPAP	YES	37.5	145\80 TO116\66	
IC30*	M	34	Jenin	16\3\2021	60%	NRFM 15L to 40 HFNC to ETT tube	Yes, Respiratory failure, Septic shock, and multiple organ failure	39.5	119\88 TO 105\58 TO 70\50	Rapid
IC32	M	48	Alaskary	5\4\2021	90%	12 L – 15 L NRM	Yes	No	125\75 TO 117\57	Slow
IC36*	M	53	Jenin	15\3\2021	82%	15 L NRM - BIPAP 15L	Yes	40.4	124\70	Rapid
IC34	F	37	PMC	21\2\2021	85%	HF NRM - HFNC	Yes	39.6	110\70	Slow
IC14	F	44	Yatta	20\2\2021	86%	CPAP	Yes	NO	120\60	
IC51	F	44	Alaskary	27\8\2021	80%	7 L FM – 15 L NRM - HFNC – 40 L CPAP	YES	NO	145\99 TO 108\60 TO 120\82	Normal
IC53*	F	46	Alaskary	23\8\2021	81%	N.C 15 L - CPAP	YES	38.5	120\70	
IC54*	M	35	Alaskary	28\8\2021	79%	5 L NC - 10 L FM – (15-20) NRM - CPAP	YES	YES	125\89-167\114-79\60	Rapid
IC55*	F	45	Jenin	9\9\2021	85%	7 L FM – (15-20) L NRM - EPAP	YES	PERSISTENT	163\76 - 109\72 - 90\75	Rapid
IC56	M	36	PMC	24\2\2021	92%	15 L NRM - HF	NO	YES	124\74 TO 107\60	Rapid
IC01	M	27	Alia	8\3\2021	80%	15 L NRM - HF	NO	38.5	121\84 TO 103\60	Rapid
IC07	M	37	Alia	21\2\2021	87%	5 L N.C - NRM HF	NO	40	117\70	

pt. code	Sodium	potassium	Bun	Albumin	creatinine	CK	Alt	Ast	LDH	Glucose
IC20	134	Normal	N	2.7	0.6	650	169	54	1000	RBS 160
IC22	143	Normal	13	3.3	0.5	22	90	58	390	FBS 140
IC23	132	5.6	17		0.76		59	57	483	93
IC24	141	4.3	29		0.93				571	RBS 212
IC49	138	low	35		0.5	143	28	17	630	RBS 117
IC26	133.5	3.7	12	4	0.86	498	33	44	434	RBS 163
IC30	144	7	139		0.6 to 7.4	322	713	2104	3385	135
IC32	130			2.9	0.54	165	57	66	770	RBS 132
IC36	136	7	99		0.9 to 6.0	766	92	65	516	85
IC34	141	3.06	11	3.57	0.5	50	80	28	791	105
IC14	127	4	11	3.5	0.5	34	19	22	340	RBS 128
IC51	144	3.3	15	3.52	0.6	274	80	39	970	
IC53	131 – 146	5.4	38	3.3	0.5	76	201	83	500	
IC54	138 - 147	5.2		1.35	0.97	200	393	168		322
IC55		5	90		0.7 to 4.8	746	90	72		199
IC56	140	4.4	21		1.1	700	93	61	321	
IC01	143	3.77	17	3.9	0.97	287	20	35.6		RBS 123
IC07	140	3.76	15		1.1		61	53	438	109

pt. code	HB	Neutrophils	lymphocyte	Platelets	aPTT	INR	WBCs	D-Dimer	Ferritin	CRP
IC20	12.5	85%	12.50%	280	44	1.4	19	188170	1860	152
IC22	12.2	62.40%	33%	212	56	1	7.7	8890	230	
IC23	11.6	88.60%	11%	541	31	1.1	6			102
IC24	12.8	92%	7.60%	359	34.4	1.9	14.8			
IC49	8.9	91%	9%	517	68	1.3	17	7042	113	
IC26	14	84.60%	15%	194	39		32	400	1262	77
IC30	15.4	84.80%	6.90%	75	102	3.56	18.8	9400	2000	
IC32	14.1	91.50%	8.30%	133			11.9	7420		198
IC36	6	89%	10%	160	48	1.2	14	7838	2000	266
IC34	12.4	85%	11.00%	208	42.6	1.14	7.5	1338	480	200
IC14	11.1	83%	10%	190	31		12.4		240	
IC51	11.5	80%	15.00%	225	28.3	1.13	17.9	7800	1580	102
IC53	10	92%	7%	114			28	2620	893	58
IC54	5.8			50			39.7	3020	6000	12.7

IC55	7.9	86%	4.6	90	180		6.5	9800		146
IC56	16	90%	10%	130	35	1.3	17.6	1862	542	
IC01	14.1	84%	14.10%	267			10.2	675		163.7
IC07	14.4	77.00%			35.9		6		970	93

Patient codes in red font represent deceased patients, while codes in blue font indicate patients who received high-flow oxygen therapy. Results highlighted in red denote abnormal values.

Appendix B: DNA Concentrations.

Sample code	Concentration ng /µL	260/280	260/230	Volume in µL	H2O µL for 30 µL Volume
IC20	300.3	1.87	2.25	1.66500	28.33500
IC22	71	1.9	2.18	7.04225	22.95775
IC23	64	1.89	2.05	7.81250	22.18750
IC24	148	1.89	2.2	3.37838	26.62162
IC26	182	1.88	2.26	2.74725	27.25275
IC01	243.4	1.9	2.1	2.05423	27.94577
IC07	372.8	1.89	2.06	1.34120	28.65880
IC14	173.7	1.88	2.07	2.87853	27.12147
IC30	216	1.89	2.23	2.31481	27.68519
IC49	25.5	1.95	1.35	19.60784	10.39216
IC32	300	1.88	2.24	1.66667	28.33333
IC34	120	1.9	2.07	4.16667	25.83333
IC51	139	1.8	1.3	3.59712	26.40288
IC53	149	1.83	1.71	3.35570	26.64430
IC54	109	1.6	0.73	4.58716	25.41284

IC55	153	1.9	2.09	3.26797	26.73203
IC56	347	1.86	1.92	1.44092	28.55908
IC36	92	1.89	1.73	5.43478	24.56522

Appendix C: Spike protein HLA Epitopes.

Allele	Description	Starting Position	Ending Position	Allele	Description	Starting Position	Ending Position
A*03:01	RLFRKSNLK	454	462	B*07:02	LPPAYTNSF	24	32
A*03:01	TVYDPLQPELDSFK	1136	1149	B*07:02	LPQGFSAL	216	223
A*03:01	VTYVPAQEK	1065	1073	B*07:02	TPINLVRDL	208	216
A*03:01	KCYGVSPTK	378	386	B*07:02	SIIAYTMSL	691	699
A*03:01	GVYFASTEK	89	97	B*07:02	YLQPRTFLL	278	286
A*03:01	GVYYHKNNK	142	150	B*07:02	EPVLKGVKL	1262	1270
A*03:01	ALDPLSETK	292	300	B*07:02	SPRRARSVA	680	688
A*03:01	EILPVSMTK	725	733	B*07:02	MIAQYTSAL	869	877
A*03:01	GVYYPDKVFR	35	44	B*07:02	FPQSAPHGV	1052	1060
A*03:01	KVFRSSVLH	41	49	B*07:02	FPREGVFV	1089	1096
A*03:01	RASANLAATK	1019	1028	B*07:02	IPTNFTISV	714	722
A*03:01	SVYAWNRKR	349	357	B*07:02	KPFERDISTEI	462	472
A*03:01	TLADAGFIK	827	835	B*07:02	LPFNDGVYF	84	92
A*24:02	PYRVVVLSF	507	515	B*07:02	LPIGINITRF	229	238
A*24:02	VYDPLQPEL	1137	1145	B*07:02	QPTESIVRF	321	329
A*24:02	GTITSGWTF	880	888	B*07:02	QPYRVVVL	506	513
A*24:02	YEQYIKWPWYI	1206	1216	B*07:02	SPRRARSV	680	687
A*24:02	YFPLQSYGF	489	497	B*07:02	TPCSFGGVSV	588	597
A*24:02	RFDNPVLPF	78	86	B*07:02	APHGVVFL	1056	1063
A*24:02	YLQPRTFLL	269	277	B*07:02	HADQLTPTW	625	633
A*24:02	GKYEQYIKW	1204	1212	B*07:02	KNIDGYFKIY	195	204
A*24:02	KYEQYIKWP	1205	1213	B*07:02	LPQGFSALEPL	216	226

A*24:02	QYIKWPWYI	1208	1216	B*07:02	NATNVVIKV	122	130
A*24:02	VYAWNRKRI	350	358	B*07:02	RVQPTESIVRF	319	329
A*24:02	KWPWYIWLGF	1211	1220	B*07:02	VGYLQPRTF	267	275
A*24:02	YEQYIKWPW	1206	1214	B*07:02	APHGVVFLH	1056	1064
A*24:02	NYNYLYRLF	448	456	B*07:02	EILDITPCSFG	583	593
A*24:02	VYSTGSNVF	635	643	B*07:02	QLTPTWRVY + PYRE(Q1)	628	636
A*24:02	AYSNNSIAI	706	714	B*07:02	WPWYIWLGF	1212	1220
A*24:02	EYVSQPFLM	169	177	B*07:02	QSAPHGVVFL	1054	1063
A*24:02	FAMQMAYRF	898	906	B*07:02	APATVCGPK	520	528
A*24:02	GYLQPRTFLL	268	277	B*07:02	APGQTGVIA		
A*24:02	HWFVTQRNF	1101	1109	B*07:02	FPNITNLCP	329	337
A*24:02	IYQTSNFRV	312	320	B*07:02	IPTDFTISV		
A*24:02	RFPNITNLCPF	328	338	B*07:02	LPPANTNSF	24	32
A*24:02	RVYSTGSNVF	634	643	B*07:02	LPPSYTNSF	24	32
A*24:02	SFPQSAPHGVVF	1051	1062	B*07:02	LPPVYTNSF	24	32
A*24:02	VFKNIDGYF	193	201	B*07:02	WPLVSSQCV		
A*24:02	VFVSNGTHWF	1094	1103	B*07:02	YLRPRTFLL	269	277
A*24:02	VYDPLQPELDSF	1137	1148	B*44:03	AEVQIDRLI	989	997
A*24:02	VYSSANNCTF	159	168	B*44:03	KEIDRLNEV	1181	1189
A*24:02	VYYPDKVF	36	43	B*44:03	REGVFVSNGTHW	1091	1102
A*24:02	YYHKNNKSW	144	152	B*44:03	TEKSNIIRGW	95	104
A*24:02	YYVGYLQPRTF	265	275	B*44:03	YEQYIKWPW	1206	1214
A*24:02	GNYNYLYRLF	447	456	B*44:03	AEIRASANL	1016	1024
A*24:02	TQDLFLPFF	51	59	B*44:03	ADAGFIKQY	829	837
A*24:02	IYKTPPIKDF	788	797	B*44:03	AEHVNNSY	653	660
A*24:02	LYNSASFSTF	368	377	B*44:03	FERDISTEIY	464	473
A*24:02	PFFSNVTWF	57	65	B*44:03	GEVFNATRF	339	347
A*24:02	SFIEDLLF	816	823	B*44:03	QELGKYEQY	1201	1209
A*24:02	SWMESEFRV	151	159	B*44:03	QEVFAQVKQIY	779	789
A*24:02	TFEYVSQPFLM	167	177	B*44:03	SETKCTLKSF	297	306

A*24:02	TYVPAQEKNFT	1066	1076	B*44:03	TECSNLLLQY	747	756
A*24:02	VFVSNGTHW	1094	1102	A*29:01	YFPLQSYGF	489	497
A*24:02	VGYLQPRTFL	267	276	B*57:03	FAMQMAYRF	898	906
A*24:02	IWLGFIAGL	1216	1224	A*26:01	CVADYSVLY	361	369
A*24:02	GYLQPRTFL	268	276	A*26:01	EILDITPCSF	583	592
A*24:02	NYNYRYRLF	448	456	A*26:01	FVFKNIDGY	192	200
A*24:02	CFTNVYADSF	391	400	A*26:01	TSNQVAVLY	604	612
A*24:02	GKYEQYIKWPWYIWL	1204	1218	A*26:01	DSKVGGNYNY	442	451
A*24:02	IKWPWYIWL	1210	1218	A*26:01	ETKCTLKSF	298	306
A*24:02	KWPWYIWLG	1211	1219	A*26:01	EVFAQVKQIY	780	789
A*24:02	LGFIAGLIA	1218	1226	A*26:01	EVFNATRFASVY	340	351
A*24:02	LQELGKYEQYIKWPW	1200	1214	A*26:01	FTISVTTEI	718	726
A*24:02	PWYIWLGFI	1213	1221	A*26:01	FVSNGTHWF	1095	1103
A*24:02	QYIKWPWYIWLGFIA	1208	1222	A*26:01	NSFTRGVYY	30	38
A*24:02	WLGFIAGLI	1217	1225	A*26:01	NTSNQVAVLY	603	612
A*24:02	WPWYIWLGF	1212	1220	A*26:01	SVASQSIIAY	686	695
A*24:02	WPWYIWLGFIAGLIA	1212	1226	A*26:01	WTAGAAAYY	258	266
A*24:02	WYIWLGFIA	1214	1222	A*26:01	WTFGAGAAL	886	894
A*24:02	YIKWPWYIW	1209	1217	A*26:01	YTMSLGAENSVAY	695	707
A*24:02	YIWLGFIAG	1215	1223	A*26:01	YTNSFTRGVY	28	37
A*24:02	LFFSDVTWF			A*30:02	QLTPTWRVY	628	636
A*24:02	QPYRVVVLS	506	514	B*53:01	LPPAYTNSF	24	32
A*24:02	VFLVLWPLV			B*53:01	FLPFFSNVTW	55	64
B*35:01	FPQSAPHGVVF	1052	1062	B*53:01	FPQSAPHGV	1052	1060
B*35:01	LPPAYTNSF	24	32	B*53:01	IAIPTNFTI	712	720
B*35:01	TSNQVAVLY	604	612	B*53:01	IHADQLTPTW	624	633
B*35:01	LTDEMIAQY	865	873	C*06:02	QYIKWPWYI	1208	1216
B*35:01	FAMQMAYRF	898	906	C*06:02	VGYQPYRVV	503	511
B*35:01	FDNPVLPFNDGVYF	79	92	C*06:02	YQPYRVVVL	505	513
B*35:01	FVSNGTHWF	1095	1103	A*02:01	ALNTLVKQL	958	966
B*35:01	IPFAMQMAY	896	904	A*02:01	FIAGLIAIV	1220	1228

B*35:01	LGAENSVAY	699	707	A*02:01	IITTDNTFV	1114	1122
B*35:01	LPFNDGVYF	84	92	A*02:01	KLPDDFMGC		
B*35:01	LPIGINITRF	229	238	A*02:01	LITGRLQSL	997	1005
B*35:01	LPPLLTDEM	861	869	A*02:01	LLFNKVTLA	821	829
B*35:01	NATRFASVY	343	351	A*02:01	NLNESLIDL	1193	1201
B*35:01	QIPFAMQMAY	895	904	A*02:01	RLDKVEAEV	983	991
B*35:01	QPTESIVRF	321	329	A*02:01	RLNEVAKNL	1185	1193
B*35:01	SANNCTFEY	162	170	A*02:01	RLQSLQTYV	1001	1009
B*35:01	VASQSIIAY	687	695	A*02:01	SIVAYTMSL		
B*35:01	WTAGAAAYY	258	266	A*02:01	VLNDILSRL	976	984
B*35:01	FCNDPFLGVY	135	144	A*02:01	VVFLHVTYV	1060	1068
B*27:05	GRLQSLQTY	999	1007	A*02:01	EILDITPCSF	583	592
C*07:01	VFAQVKQIY	781	789	A*02:01	KLPDDFTGCV	424	433
C*07:01	FKNLREFVF	186	194	A*02:01	LLALHRSYL	241	249
C*07:01	FRSSVLHST	43	51	A*02:01	TLDSKTQSL	109	117
C*07:01	TRFQTLLAL	236	244	A*02:01	YYVGYLQPRTFLL	265	277
B*40:01	SEPVLKGVKL	1261	1270	A*02:01	GLTVLPPLL	857	865
B*40:01	KEIDRLNEV	1181	1189	A*02:01	KLNDLCFTNV	386	395
B*40:01	NESLIDLQEL	1194	1203	A*02:01	SLIDLQEL	1196	1203
B*40:01	AEIRASANL	1016	1024	A*02:01	SIIAYTMSL	691	699
B*40:01	AEVQIDRL	989	996	A*02:01	YLQPRTFLL	269	277
B*40:01	FDEDDSEPVL	1256	1265	A*02:01	MIAQYTSAL	869	877
B*40:01	FERDISTEI	464	472	A*02:01	FTISVTTEI	718	726
B*40:01	FEYVSQPFLM	168	177	A*02:01	HLMSFPQSA	1048	1056
B*40:01	GEVFNATRF	339	347	A*02:01	KIADYNYKL	417	425
B*40:01	TESIVRFPNITNL	323	335	A*02:01	YGFQPTNGV	495	503
B*40:01	YECDIPIGAGI	660	670	A*02:01	YQDVNCTEV	612	620
A*01:01	CVADYSVLY	361	369	A*02:01	FLPFFSNV	55	62
A*01:01	FCNDPFLGVYY	135	145	A*02:01	CNDPFLGVY	134	142
A*01:01	STQDLFLPFF	50	59	A*02:01	ELLHAPATV	516	524
A*01:01	TSNQVAVLY	604	612	A*02:01	FELLHAPATV	515	524
A*01:01	LTDEMIAQY	865	873	A*02:01	FVFLVLLPL	2	10
A*01:01	SANNCTFEY	162	170	A*02:01	HADQLTPTW	625	633
A*01:01	WTAGAAAYY	258	266	A*02:01	KNIDGYFKIY	195	204

A*01:01	YTNSFTRGVY	28	37	A*02:01	NATNVVIKV	122	130
A*01:01	LADAGFIKQY	828	837	A*02:01	RLDKVEAEVQI	983	993
A*01:01	CNDPFLGVY	136	144	A*02:01	SFELLHAPATV	514	524
A*01:01	ITGRLQSLQTY	997	1007	A*02:01	SWMESEFRV	151	159
A*01:01	LLTDEMIAQY	864	873	A*02:01	VGYLQPRTF	267	275
A*01:01	LQELGKYEQY	1200	1209	A*02:01	VLPFNDGVYFA	83	93
A*01:01	LTDEMIAQYT	865	874	A*02:01	YFQPRTFLL		
A*01:01	SASFSTFKCY	371	380	A*02:01	FQFCNDPFL	133	141
A*01:01	SKRVDFCGKGY	1037	1047	A*02:01	SYQTQTNSPRRA	673	684
A*01:01	SSANNCTFEY	161	170	A*02:01	VLYENQKLI	915	923
A*01:01	STECSNLLLQY	746	756	A*02:01	YLQLRTFLL		
A*01:01	STQDLFLPF	50	58	A*02:01	YLQPRIFLL		
A*01:01	TDEMIAQY	866	873	A*02:01	NIADYNYKL	414	422
A*01:01	YTNSFTRGVYY	28	38	A*02:01	TIADYNYKL	417	425
A*01:01	TSNEVAVLY			A*02:01	FLHVTYVPA	1062	1070
A*01:01	FSDVTWFHA			A*02:01	KLPDDFTGC	424	432
A*01:01	FSNVTWFHA	59	67	A*02:01	KQIYKTPPI	786	794
A*01:01	NSAIGKIQY	931	939	A*02:01	KQLSSNFGA	964	972
A*01:01	NSTRFASVY	343	351	A*02:01	SVTTEILPV	721	729
A*01:01	STECSNLLL	746	754	A*02:01	VTWFHAIHV	62	70
A*01:01	VSDGTHWFV			A*02:01	EILDITPCSFG	583	593
A*01:01	WMESEFRVY	152	160	A*02:01	QLTPTWRVY + PYRE(Q1)	628	636
A*01:01	YSSANNCTF	160	168	A*02:01	RLITGRLQSL	995	1004
B*57:01	GTITSGWTF	880	888	A*02:01	FVFLVLLPLV	2	11
B*57:01	TSNQVAVLY	604	612	A*02:01	RLQSLQIYV	1000	1008
B*57:01	GVFVSNGTHW	1093	1102	A*02:01	GQTGKIADYNYKL	413	425
B*57:01	NSIAIPTNF	710	718	A*02:01	YLQPRTFLLKYNE	269	281
B*57:01	RSVASQSII	685	693	A*02:01	AIPINFTISV	710	719
B*57:01	VAIHADQLTPTW	622	633	A*02:01	AIPTNFTISV	713	722
C*07:02	TLDSKTQSL	109	117	A*02:01	CNDPFLGVYY	136	145
C*07:02	RFDNPVLPF	78	86	A*02:01	ELDSFKEEL	1144	1152

C*07:02	CTFEYVSQPFLMDLE	175	189	A*02:01	FCNDPFLGV	135	143
C*07:02	LTDEMIAQY	865	873	A*02:01	FCNDPFLGVY	133	142
C*07:02	IYKTPPIKDF	788	797	A*02:01	RDIADTTDAV	567	576
C*07:02	QSAPHGVVF	1054	1062	A*02:01	RVVVLSFEL	509	517
C*07:02	TRFASVYAW	345	353	A*02:01	SRLDKVEAEV	982	991
C*07:02	TRTQLPPAY	20	28	A*02:01	VLNDILARL	973	981
A*30:01	VTYVPAQEK	1065	1073	A*02:01	VTWFHAISG	62	70
A*30:01	KVFRSSVLH	41	49	A*02:01	FVFFVLLPLV	2	11
A*30:01	RVYSTGSNV	634	642	A*02:01	YIWLGFIAGL	1215	1224
A*30:01	ASVYAWNRK	348	356	A*02:01	FLPFFSDVT		
A*30:01	ATRFASVYA	344	352	A*02:01	FLPFFSNVT	55	63
A*30:01	KNLREFVFK	187	195	A*02:01	FLVLLPLVS	4	12
B*51:01	DAVRDPQTL	574	582	A*02:01	GLIAIVMVT	1223	1231
B*51:01	EPLVDLPI	224	231	A*02:01	ILPDPSKPS	805	813
B*51:01	EVFAQVKQI	780	788	A*02:01	IMLCCMTSC	1232	1240
B*51:01	FAMQMAYRF	898	906	A*02:01	ITSGWTFGA	882	890
B*51:01	FPREGVFV	1089	1096	A*02:01	LPDDFMGCV		
B*51:01	IAIPTNFTI	712	720	A*02:01	LPDDFTGCV	425	433
B*51:01	IANQFNSAI	923	931	A*02:01	NLDESLIDL		
B*51:01	IPFAMQMAY	896	904	A*02:01	SLSSTASAL	937	945
B*51:01	IPTNFTISV	714	722	A*02:01	VFLVLLPLV	3	11
B*51:01	LPFNDGVYF	84	92	A*02:01	VLSFELLHA	512	520
B*51:01	LPLVSSQCV	8	16	A*02:01	YQDVDCTEV		
B*51:01	LPPLLTDEM	861	869	A*02:01	YTNSFTRGV	28	36
B*51:01	YAWNRKRI	351	358	A*11:01	ASANLAATK	1020	1028
B*51:01	YGFQPTNGV	495	503	A*11:01	SLIDLQELGK	1196	1205
B*51:01	NATNVVIKV	122	130	A*11:01	GTHWFVTQR	1099	1107
B*51:01	VGYLQPRTF	267	275	A*11:01	RLFRKSNLK	454	462
B*51:01	FASTEKSNI	92	100	A*11:01	VTYVPAQEK	1065	1073
C*15:02	VVNQNAQAL	951	959	A*11:01	KCYGVSPTK	378	386
A*68:01	SVLNDILSR	975	983	A*11:01	GVYFASTEK	89	97
A*68:01	GTHWFVTQR	1099	1107	A*11:01	GVYYHKNNK	142	150
A*68:01	GVYFASTEK	89	97	A*11:01	NSASFSTFK	370	378
A*68:01	NSASFSTFK	370	378	A*11:01	TLKSFTVEK	302	310
A*68:01	DGVYFASTEK	88	97	A*11:01	ASVYAWNRK	348	356
A*68:01	EILPVSMTK	725	733	A*11:01	FIEDLLFNK	817	825
A*68:01	FASVYAWNR	347	355	A*11:01	GVYHKNNK	140	147
A*68:01	FVIRGDEVR	400	408	A*11:01	CTLKSFTVEK	301	310

A*68:01	GVYYPDKVFR	35	44	B*08:01	FPQSAPHGVVF	1052	1062
A*68:01	HVSGTNGTK	69	77	B*08:01	LPQGFSAL	216	223
A*68:01	HVTYVPAQEK	1064	1073	B*08:01	TLDSKTQSL	109	117
A*68:01	NASVVNIQK	1173	1181	B*08:01	SIIAYTMSL	691	699
A*68:01	NSFTRGVYY	30	38	B*08:01	YLQPRTFLL	269	277
A*68:01	NVYADSFVIR	394	403	B*08:01	MIAQYTSAL	869	877
A*68:01	QIAPGQTGK	409	417	B*08:01	INITRFQTL	233	241
A*68:01	STGSNVFQTR	637	646	B*08:01	KIYSKHTPI	202	210
A*23:01	QYIKWPWYI	1208	1216	B*08:01	NITRFQTL	234	241
A*23:01	QYIKWPWYIW	1208	1217	B*08:01	QPYRVVVL	506	513
B*07:02	APHGVVFLHV	1056	1065	B*08:01	SPRRARSV	680	687
B*07:02	QPYRVVVLSF	506	515	C*12:03	FVSNGTHWF	1095	1103
B*07:02	EILDITPCSF	583	592				

Appendix D: SARS COV2 Nucleocapsid protein epitopes.

Allele Name	Epitope Name	Starting Position	Ending Position	Allele Name	Epitope Name	Starting Position	Ending Position
A*03:01	KTFPPTEPK	361	369	B*35:01	LLNKHIDAY	352	360
A*03:01	KTFPPTEPKK	361	370	B*35:01	SPDDQIGYY	79	87
A*03:01	LLNKHIDAYKTFPPTEPK	352	369	B*35:01	QFAPSASAF	306	314
A*03:01	QLPQGTTLPK	160	169	B*35:01	SSPDDQIGYY	78	87
A*03:01	KTFPPTEPKKDKKK	361	374	C*07:02	LSPRWYFYYLGTGPEAGL	104	121
B*27:05	MSDNGPQNQRNAPRITF	1	17	C*07:02	QRNAPRITF	9	17
B*27:05	NQRNAPRITFGGPSDSTG	8	25	C*07:01	LLLDRLNQL	222	230
B*27:05	QRNAPRITF	9	17	C*07:01	FAPSASAFF	307	315
B*35:01	FPRGQGVPI	66	74	C*07:01	QRNAPRITF	9	17
B*35:01	TPSGTWLTY	325	333	C*07:01	SPDDQIGYY	79	87
B*35:01	FAPSASAFF	307	315	C*07:01	KKQQTVTLL	387	395
B*35:01	KAYNVTQAF	266	274	C*07:01	MKDLSPRWY	101	109
B*35:01	LPAADLDDF	395	403	C*07:01	RRGPEQTQGNF	276	286

B*35:01	LPNNTASWF	45	53				

Appendix E: Python Code Snippet

```
import pandas as pd
import matplotlib.pyplot as plt
import pathlib

font = {'family' : 'normal',
        'weight' : 'normal',
        'size'   : 15}

plt.rc('font', **font)

# use textwrap from python standard lib to help manage how the description
# text shows up
import textwrap

# load best results without patient
df = pd.read_excel('Epitops_Data.xlsx',sheet_name='epitops-TABLE 3.3.1')
#df = pd.read_excel('Epitops_Data.xlsx',sheet_name='Signficant_HLA')

df.head()

df = df.sort_values(by='Allele', ascending=True).reset_index(drop=True)
df.head()
```

	Patiant	Allele	Epitops	Starting_Position	Ending_Position	Score
0		B*57:03	FAMQMAYRF	898	906	NaN
1		C*07:01	VFAQVKQIY	781	789	NaN
2		C*07:01	FKNLREFVF	186	194	NaN
3		C*07:01	FRSSVLHST	43	51	NaN
4		C*07:01	TRFQTLLAL	236	244	NaN

Python code: https://drive.google.com/file/d/1T6xv_VJdYLZsxmDKiUFQfkG-Hzuqf_Ku/view?usp=sharing

Epitopes Dataset: https://docs.google.com/spreadsheets/d/1HCok3XrAp78EOCrHmr--a-XkvPiRJ-DQ/edit?usp=sharing&ouid=113042661501879228518&rtpof=true&sd=true

Appendix F: IRB Approval

المجلس الفلسطيني للبحث الصحي

Palestinian Health Research Council

تعزيز النظام الصحي الفلسطيني من خلال مأسسة استخدام المعلومات البحثية في صنع القرار

Developing the Palestinian health system through institutionalizing the use of information in decision making

Helsinki Committee

For Ethical Approval

Date: 05\04\2021 **Number:** PHRC/HC/872/21

Name: Dr Nouar Qutob الاسم:

We would like to inform you that the committee had discussed the proposal of your study about: نفيدكم علماً بأن اللجنة قد ناقشت مقترح دراستكم حول:

Correlation between HLA type and Covid 19 severity among the Palestinian population

The committee has decided to approve the above mentioned research. Approval number PHRC/HC/872/21 in its meeting on 05\04\2021 و قد قررت الموافقة على البحث المذكور عاليه بالرقم والتاريخ المذكوران عاليه

Signature

Member Dr. Khamis Elkess

Member Dr. Yehia Abed

Chairman Nasser D. Abu Shanab

Genral Conditions:-

1. Valid for 2 years from the date of approval.
2. It is necessary to notify the committee of any change in the approved study protocol.
3. The committee appreciates receiving a copy of your final research when completed.

Specific Conditions:-

E-Mail:pal.phrc@gmail.com

Gaza - Palestine غزة - فلسطين

شارع النصر - مفترق العيون

العلاقة بين نظام مستضدات الكريات البيض البشرية (HLA) و حدة مرض كوفيد 19 في الشعب الفلسطيني

منى مفيد يوسف محمود

أسماء لجنة الإشراف: الدكتور زيدون صلاح

الدكتورة نوار قطب

الدكتور روبين أبو غزالة

الدكتور كمال الضميدي

ملخص

أعلنت منظمة الصـــحة العالمية أن مرض فيروس كورونا 19، الناجم عن SARS-CoV-2 جائحة عالمية في 11 مارس 2020. وقد بلغ معدل الوفيات 5708، مع أكثر من 703228 حالة إصـابة بهذا الوبـاء في الضـفة الغربية و قطاع غزة . يتباين تطور الأعراض المصـاحبة للمرض بشـــكل كبير، بدءًا من الحالات التي لا تظهر عليها أعراض إلى الحالات الشـــديدة والحرجة المميتة. أصـبح هذا التباين في مدى حدة المرض نقطة محورية للدراسـات البحثية في جميع أنحاء العالم. وقد أشار العلماء إلى دور آلية تقديم المستضـد –نظام مستضـدات الكريات البيض البشـرية (HLA)– في شـدة مرض كوفيد 19. لذلك كان الهدف من هذا البحث هو دراسـة الارتباطات الإحصـائية المحتملة بين HLA-I وشـدة مرض كوفيد–19 بين مجموعة فلسـطينية تضـم 18 مريضًـا في وحدة العناية المركزة يعانون من أمراض كوفيد–19 الشـديدة

والحرجة. ولتحقيق هذا الهدف، تم إجراء تسلسل الإكسوم الكامل (WES) باستخدام تقنية (NGS) على الحمض النووي المستخرج من دم مرضى كوفيد−19 في وحدة العناية المركزة. تم اعتماد نتائج التنميط الجيني الخاصة ببرنامج HISAT genotyping حيث تم استخدام ثلاثة نماذج برمجية خوارزمية (Seq2HLA ,HISAT genotyping ,HLA-LA) لمقارنة الدقة بينها، ثم تم تقييم الارتباط بين أليلات HLA-I والعدوى الحرجة ل لـ COVID-19 بالدلالات الإحصائيَة باستخدام برنامج Prism، وتطبيق اختباري ,Fisher's Exact) (continuity-corrected Yates χ2. تم إجراء معادلة Benjamini-Hochberg لحساب قيمة P المصححة. علاوة على ذلك، تم تطوير كود بايثون لتصورمواقع الحواتم (Epitopes) المتعلقة بال Spike protein و ال Nucleocapsid لفيروس SARS−COV 2 لأليلات HLAالمهمة الناتجة. تم استكشاف الحواتم عبر قاعدة بيانات الحواتم المناعية (IEDB). كشف تحليلنا أن بعض أليلات HLA-I مرتبطة ايجابيا بشكل كبير (P <0.05) بحالات كوفيد−19 الحرجة والشديدة، مما يشير الى دور محتمل لها في زيادة حدة المرض تتضمن هذه الأليلات كل من A*01:03، B*15:10، B*57:03، C*03:04 C*07:01، C*07:02، C*14:02 علاوة على ذلك، تشير دراستنا إلى احتمالية أن ارتباطات أليلات HLA-C*07:01 و HLA-C*07:02 مع حواتم الخلايا التائية قد تؤدي إلى تسهيل دخول الفيروس إلى الخلية البشرية المعرضة للفيروس.

كلمات مفتاحية: كوفيد−19، الضفة الغربية، نظام مستضدات الكريات البيض البشرية (HLA)، فيروس SARS- CoV-2