VDR Gene Variants and their Association with HIV Progression

Submitted by:

Kainat Mohammed Rashid 00000356922

Warda Khalid 00000356938

Supervisor:

Dr: Maria Shabbir

Co-Supervisor:

Dr: Yasmin Badshah

Atta-Ur-Rahman School of Applied Biosciences

National University of Sciences and Technology

Islamabad, Pakistan

2024

VDR Gene Variants and their Association with HIV Progression

Submitted by:

Kainat Mohammed Rashid 00000356922

Warda Khalid 00000356938

Supervisor:

Dr: Maria Shabbir

Co-Supervisor:

Dr: Yasmin Badshah

This thesis is submitted to the National University of Sciences and Technology, Islamabad, in partial fulfillment of the requirements for the degree of Bachelor of Sciences in Applied Biosciences

Atta-Ur-Rahman School of Applied Biosciences

National University of Sciences and Technology

Islamabad, Pakistan

2024

THESIS ACCEPTANCE CERTIFICATE

This document certifies that this is the final copy of BS FYP Thesis written by **Ms. Kainat Mohammed Rashid** (Reg No. 356922) and **Ms. Warda Khalid** (Reg No. 356938), of **Atta-Ur-Rahman School of Applied Biosciences (ASAB)**. The undersigned has thoroughly reviewed and found that the document complies with all regulations set by the National University of Sciences and Technology (NUST). It is confirmed to be free of plagiarism, errors, and mistakes, and is accepted as a partial fulfillment for the Bachelor of Sciences degree in Applied Biosciences. Additionally, all necessary amendments suggested during the final presentation have been included in the thesis.

Signature: _____

Dr. Maria Shabbir
Tenured Associate Professor
Healthcare Biotechnology
Atta-ur-Rahman School of Applied
Biosciences (ASAB), NUST Islamabad

Name of Supervisor: **Dr. Maria Shabbir**

Date: 30-05-24

Signature: _____

Dr. Yasmin Badshah
Assistant Professor
Atta-ur-Rahman School of
Applied Biosciences (ASAB)
NUST Islamabad

Name of Co-Supervisor: **Dr. Yasmin Badshah**

Date: 30-05-24

Signature (HOD): _____

Dr. Hussain Mustatab Wahedi
Head of Department (HoD)
Dept of Healthcare Biotechnology
Atta-ur-Rahman School of Applied
Biosciences (ASAB), NUST Islamabad

Head of Department Biomedicine: **Dr. Hussain Mustatab Wahedi**

Date: 31/05/2024

Signature (Dean/Principal): _____

Principal ASAB: **Dr. Muhammad Asghar**

Date: 31/05/24

DECLARATION

We affirm that the research work titled **" VDR Gene Variants and their Association with HIV Progression"** authored by **Kainat Mohammad Rashid** and **Warda Khalid** is our originalcreation. This work has not been presented elsewhere for evaluation. The research was conducted during our undergraduate studies at Atta-ur-Rahman School of Applied Biosciences (ASAB), National University of Sciences and Technology (NUST), under the guidance of Dr. Maria Shabbir. Any material borrowed from external sources has been appropriately cited and acknowledged.

Kainat Mohammad Rashid

Warda Khalid

CERTIFICATE FOR PLAGIARISM

It is certified that this Undergraduate Thesis Titled **"VDR Gene Variants and their Association with HIV Progression"** written by **Ms. Kainat Mohammed Rashid** (Reg No. 356922) and **Ms. Warda Khalid** (Reg No. 356938) of **Atta-Ur-Rahman School of Applied Biosciences (ASAB),** has been examined by me.

I undertake that:

a. The thesis contains original contributions and novel insights not found in previously published works or those currently being reviewed for publication elsewhere. No exact sentences, equations, diagrams, tables, paragraphs, or sections have been directly copied from prior sources without proper quotation marks and citations.

b. The work presented is original and own work of the authors i.e., there is no plagiarism.

c. There is no falsification of data or misrepresentation of research findings in the records. The thesis has undergone inspection using TURNITIN (an originality report is attached), confirming compliance with the HEC plagiarism policy and relevant guidelines.

Dr. Maria Shabbir
Tenured Associate Professor
Deptt of Healthcare Biotechnology
Atta-ur-Rahman School of Applied Biosciences (ASAB), NUST Islamabad

Signature of Supervisor

Dr. Maria Shabbir

Dr. Yasmin Badshah
Assistant Professor
Atta-ur-Rahman School of Applied Biosciences (ASAB)
NUST Islamabad

Signature of Co-Supervisor

Dr. Yasmin Badshah

Student Signatures

Kainat Mohammad Rashid:

Hansa

Warda Khalid:

Warda

"Do not falter or grieve, for you will have the upper hand, if you are ˹true˺ believers."

Quran 3:139

DEDICATION

"This thesis is dedicated to our parents, siblings, and teachers for their unwavering support, encouragement, and guidance throughout our journey. Thank you for believing in us".

Acknowledgments

I am profoundly thankful to Allah Almighty for blessing me with the opportunity and resources to undertake this research. I owe a deep sense of gratitude to **Dr. Muhammad Asghar, Principal** of the Atta ur Rahman School of Applied Biosciences at the National University of Science and Technology, Islamabad, for generously providing access to the necessary facilities. I am equally thankful to **Dr. Hussain Mustatab Wahedi**, Head of the Department, for his support throughout this journey. My heartfelt thanks go to my supervisor **Dr. Maria Shabbir**, Associate Professor, for her invaluable guidance and persistent encouragement which were crucial for the successful completion of this project. Her detailed feedback greatly improved my manuscript. I am also grateful to my co-supervisor **Dr. Yasmin Badshah**, Assistant Professor, for her guidance and motivational support during my research. Additionally, my sincere appreciation extends to all the ASAB staff and faculty whose support and cooperation have been instrumental in my academic journey. We are grateful to our seniors, **Ms. Amna Hafeez, Ms. Aneela Mustafa** and **Ms. Sameen Zafar** for their guidance and support in this project.

Table of Contents

List of Figures:

List of Tables:

List of Abbreviations

AIDS - Acquired Immunodeficiency Syndrome.

ARMS PCR - Allele-Specific PCR Amplification Refractory Mutation System Polymerase Chain Reaction

ART - Antiretroviral Therapy

CADD -Combined Annotation Dependent Depletion

CBP - CREB Binding Protein

CD4 - Cluster of Differentiation 4

CI - Confidence Interval

DBP - Vitamin D Binding Protein

HIV -Human Immunodeficiency Virus

LTR - Long Terminal Repeat

MetaLR - Meta Logistic Regression

NAT - Nucleic Acid Test

OR - Odds Ratio

PCR - Polymerase Chain Reaction

PDB - Protein Data Bank

POCT - Point-of-Care Testing

PolyPhen -Polymorphism Phenotyping

PrEP - Pre-exposure Prophylaxis.

Revel- Rare Exome Variant Ensemble Learner

RR - Relative Risk

RXR - Retinoid X Receptor

SDG -Sustainable Development Goal

SDS - Sodium Dodecyl Sulfate

SIFT -Sorting Intolerant from Tolerant

SNP - Single Nucleotide Polymorphism

SRC - Steroid Receptor Coactivator

UV - Ultraviolet

VDR - Vitamin D Receptor

VDRE - Vitamin D Response Element

WHO - World Health Organization

Abstract

With HIV (Human Immunodeficiency Virus) posing a continuous global challenge, our study aimed to investigate the diagnostic potential of the VDR gene variant rs746619116. Through studying over 28,000 variants, we discovered three significant pathogenic missense mutations from which rs746619116 was selected. The rs746619116 replaces arginine with tryptophan at position 154. Using Project HOPE and DynaMut2, we found that this mutation leads to important structural disruptions in the VDR protein, reducing its stability and functionality. Genetic studies revealed a higher incidence vulnerability of the CT genotype among HIV-positive patients, indicating a possible increase in susceptibility. Moreover, this genotype was associated with higher vitamin D levels, platelet counts, and an elevated risk of thalassemia. Although promising, these findings require further validation with more extensive and diverse populations.

Chapter 1: Introduction

HIV, short for human immunodeficiency virus, is responsible for causing HIV infection, a condition where the virus attacks and weakens the body's infection-fighting CD4 cells, also known as CD4 T lymphocytes. This weakening of the immune system makes it challenging for the body to defend against infections, illnesses, and specific types of cancer. If left untreated, HIV can progressively damage the immune system further, leading to a decline in health and eventually the development of AIDS, or acquired immunodeficiency syndrome. However, with proper treatment, the immune system can regain strength and functionality (HIV and AIDS, 2023). HIV, a retrovirus that infects humans, targets specific white blood cells known as CD4 T cells, crucial components of the immune system responsible for fighting infections. However, HIV manipulates these cells, causing them to replicate the virus while gradually damaging and reducing the T cell count. In a healthy individual, there are typically between 750 and 1,500 CD4 T cells per microliter of blood. As HIV progresses, the number of these cells declines over time, compromising the immune system's ability to combat infections. When individuals with HIV have fewer than 200 CD4 T cells per microliter of blood, they become susceptible to severe illnesses or infections, indicating the progression to AIDS (Organization, HIV/AIDS, 2024). Since the onset of the HIV epidemic, approximately 85.6 million people have been infected with the virus, and tragically, around 40.4 million have lost their lives due to HIV-related causes. Currently, an estimated 39.0 million individuals are living with HIV globally, with approximately 0.7% of adults aged 15–49 years affected. The impact of the epidemic varies significantly across countries and regions. The WHO African Region bears the greatest burden, with nearly 3.2% of adults living with HIV, constituting more than two-thirds of the global HIV-positive population. In 2022, approximately 630,000 individuals died from HIV-related illnesses worldwide. Prevention efforts have been notable, with 90 low- and middle-income countries reporting a total of 190 million people tested and received results in 2018, showcasing ongoing efforts to combat the spread of HIV (Organization, HIV, 2022).

1.1 Stages of HIV

During the initial acute infection phase lasting 2 to 10 weeks, there is a notable decline in CD4+T cell levels along with a significant increase in circulating free virus. This phase manifests with acute symptoms like fever, lymphadenopathy, pharyngitis, headache, and rash. After this acute phase,

CD4+T cell levels return to almost normal, and the viral load decreases significantly. Despite continuous HIV replication, individuals do not display major disease symptoms during this asymptomatic or latent period, which typically spans 7 to 10 years. Following the asymptomatic phase there is a swift increase in viral load coupled with a decline in CD4+T cell count. This advanced stage, known as AIDS, is identified by CD4+T cell counts dropping below 200 cells/mm3 (Olga Shcherbatova, 2020).

1.2 Transmission

HIV, transmitted via bodily fluids like blood, semen, vaginal fluids, or breast milk, spreads within the body through target cells as either free viral particles or cell-associated forms. In cell-associated spread, the virus rapidly enters target cells at points of cell-to-cell contact, offering partial protection from the external environment and concentrating viral particles at infection sites. Conversely, in cell-free spread, replicated viruses diffuse and attach to CD4 receptors on target cells. While lab studies suggest cell-associated spread is more effective than cell-free spread, opinions on their in vivo efficiency vary. The origins of most HIV transmissions, notably male-to-female and mother-to-child routes, remain uncertain regarding cell-free or cell-associated virus involvement. Despite correlations between cell-free virus levels and infectiousness, transmission can occur even when the cell-free virus is undetectable, indicating a potential role for cell-associated (Sarudzai P. Showa, 2019).

1.3 Diagnosis

Currently, various methods are employed to diagnose HIV infection. These methods serve crucial roles in public health efforts, clinical settings, and individual health management. Among the primary diagnostic approaches are antibody tests, antigen/antibody tests, and nucleic acid tests (NATs). Each method has its own set of characteristics and window periods, which influence their effectiveness in detecting HIV at different stages post-exposure. (Geng, 2021) HIV antibodies can be found in blood or oral fluids using antibody tests, which typically detect the antibodies 23 to 90 days after exposure. This includes the majority of quick tests and the only HIV self-test that has received FDA approval. Interestingly, tests that take blood directly from a vein typically identify HIV earlier than those that use oral fluids or blood from a finger stick. In the meantime, antigen/antibody tests, which are frequently carried out in laboratories, search for antigens such as

the p24 antigen, which manifests early in HIV infections, as well as HIV antibodies. These tests are the accepted norm in the United States and typically detect HIV 18 to 45 days after exposure when using venous blood, and 18 to 90 days when using finger stick blood samples. Additionally, nucleic acid tests, (NATs) are especially useful for people who have recently been exposed to HIV or have previously tested negative for the virus. NATs identify HIV directly in the bloodstream. A shorter detection period is provided by NATs, which can identify HIV between 10 and 33 days after exposure (Watson, 2023).

1.4 Problems with Current HIV Diagnosis Methods

Diagnosing HIV involves significant challenges. Many prevalent HIV tests, like the antibody-based point-of-care tests (POCTs), do not detect HIV in its earliest phase when viral loads are high, but antibodies are not yet present. Starting antiretroviral therapy (ART) during this early phase can also suppress or delay the formation of HIV-specific antibodies, adding complexity to the diagnosis. This problem is particularly acute in resource-poor settings where antibody-based POCTs are widely used, which can result in cases going undiagnosed, delays in starting treatment, increased risk of transmission, and potential development of drug resistance. As the use of ART and pre-exposure prophylaxis (PrEP) expands, there is a crucial need for ongoing research to improve and develop cost-effective diagnostic techniques. The diagnostic process is further complicated by factors such as the patient's demographic background and the stage of infection. In particular, diagnosing HIV in infants under 18 months old is difficult due to the presence of maternal antibodies and the latency in the formation of detectable antibodies. (Tamara Elliott, 2019).Furthermore, the inability of current diagnostic techniques to differentiate between recent and long-lasting infections leads to diagnostic uncertainty even though this difference is vital for understanding the virus's mode of transmission. Testing is made more difficult by the diversity of HIV strains, as certain strains may not be picked up by standard methods. Suggestions for attending to these challenges include targeting universally conserved viral markers, tailoring tests to specific local areas, and advancing testing technologies to increase diagnostic accuracy.

The complexities surrounding HIV diagnosis, especially in regions like Western Africa, are influenced by various factors. The presence of diverse HIV-1 subtypes poses challenges to accurate

diagnosis, particularly within Groups M and O where notable inaccuracies are observed (Rachel S. Kamgaing, 2023). Advanced testing for HIV is often not available in low- and middle-income countries. This lack of sophisticated tests, like those that check for viral load or specific HIV DNA, can lead to incorrect positive test results. These errors can cause unneeded treatment and stress for people. Additionally, when standard tests give unclear results, other methods like trying to isolate the virus are needed, but these can be difficult if the virus levels are too low to detect.

1.5 VDR Gene and HIV diagnosis

Research indicates a link between the vitamin receptor (VDR) gene polymorphisms and susceptibility to viral infections, including HIV. For instance, a study by (Gema Nieto, 2004)highlighted the role of VDR gene variants, particularly the Fok-I polymorphism (rs2228570), in altering immune responses and increasing vulnerability to HIV infection. Furthermore, Certain variants of the VDR gene, particularly those identified as Bsm-I and Apa-I, are proposed as potential diagnostic markers for HIV infection, based on their distribution among HIV/AIDS patients in West Java and their suggested association with disease progression. (Hendro Hendro, 2020)Vitamin D plays a crucial role in immune system regulation, and deficiency of it can increase the risk to infections. The discovery of the vitamin D receptor (VDR) suggested its vital role in immune functionality, as activating VDR affects the function of genes that direct immune responses. The relationship between variations in the VDR gene and HIV diagnosis is important for developing early detection methods. Understanding the genetic vulnerability linked to VDR polymorphisms helps identify individuals who may be at an increased risk of contracting HIV or undergoing rapid disease progression. This insight allows for specific management and the timely start of treatment, considerably improving clinical outcomes and minimizing the spread of the virus within groups. Variants of the VDR gene are also being explored as potential biomarkers for the early detection of HIV, offering important hope for offering health services unique to every individual and strategic public health action. Additionally study done by (Meropi Dimitriadou, 2011) focused on analyzing the distribution of the Fok-I polymorphism within the VDR gene among Greek children and young adults diagnosed with beta-thalassemia the study concluded that individuals carrying the f allele were more prone to having lower concentrations of the vitamin D metabolite $1,25(OH)2D_3$. These results point to a possible genetic influence on vitamin D metabolism in these patients suggesting the role of this gene in different diseases.

1.6 Objective

This thesis aims to assess if the VDR gene variant rs746619116 can be used as a diagnostic marker for HIV. This genetic variant, characterized as a single nucleotide polymorphism (SNP), involves a change from G to A in the DNA sequence. It is found on chromosome 12 at position 47,857,506 in the GRCh38 genome assembly or position 48,251,289 in the GRCh37 assembly. The afflicted gene is VDR. This variant is classified as a missense variant and coding sequence variant, indicating that it alters the amino acid sequence of the encoded protein. The clinical significance of this variant is uncertain.

- This entails analyzing the variant's potential pathogenicity through in-depth in *in silico* analysis and validating these results using multiple bioinformatics tools. The main focus is to determine if rs746619116 correlates with HIV infection status.
- Confirming these findings through clinical studies testing patients diagnosed with HIV. This research aims to establish the reliability of rs746619116 as a straightforward diagnostic indicator for identifying individuals with HIV or those at risk of disease progression.

Chapter 2: Literature Review

2.1 Global HIV Trends: A Focus on Rising Cases in Pakistan

HIV (Human Immunodeficiency Virus), a persistent global health challenge, has claimed the lives of approximately 40.4 million individuals worldwide, with the majority of cases concentrated in the WHO African region. HIV is characterized by a gradual progression wherein the virus selectively targets CD4 T helper cells, progressively compromising the immune system's integrity. Recognizing the urgency of addressing this crisis, organizations like WHO, along with collaborative efforts with stakeholders, have set ambitious targets to end the HIV epidemic by 2030, aligning with the Sustainable Development Goals (SDGs). (Organization, 2024) In recent years, Pakistan has witnessed a concerning rise in HIV cases, with the first reported case dating back to 1987, stemming from unsafe blood transfusions. Subsequent outbreaks, such as the one in Larkana (Sindh region) in 2004, among injection drug users, have underscored the vulnerability of specific populations to HIV transmission. As of 2021, it is estimated that approximately 3.8 million individuals in Pakistan are living with HIV, with marginalized communities such as the intersex community and sex workers being disproportionately affected (Muhammad Aizaz, 2023).

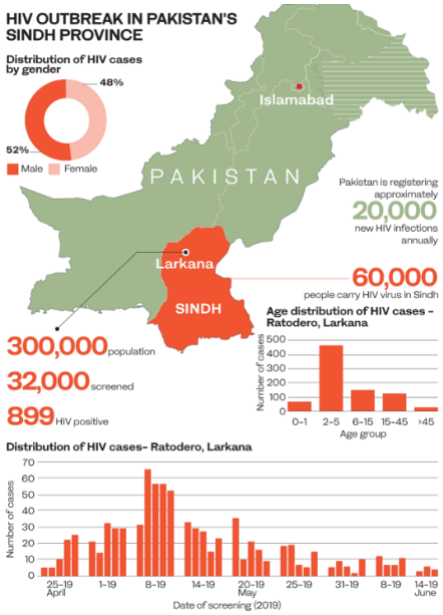

Figure 1: Pakistan's Sindh grapples with outbreak of HIV infections (Pakistan's Sindh grapples with outbreak of HIV infections, 2019).

2.2 Overview of HIV

HIV presents itself in two primary forms: HIV-1 and HIV-2. HIV-1 accounts for approximately 95% of global HIV cases and typically exhibits a slower progression and transmission rate compared to HIV-2. Conversely, HIV-2 is predominantly found in specific regions, notably Western Africa. Importantly, while medications used to treat HIV-1 may also be effective against HIV-2, they can induce resistance in HIV-2 strains, highlighting the nuanced challenges in combating different strains of the virus. (MedicalNewsToday, n.d.). Despite advancements in medical science, HIV remains incurable. However, with timely and appropriate medical intervention, the progression of the disease can be effectively managed. The origin of HIV is a subject of considerable scientific inquiry, with evidence pointing to zoonotic transmission from primates, particularly chimpanzees, to humans. Historical case studies dating back to the 1800s suggest that human exposure to chimpanzee blood during hunting practices may have initiated the cross-species transmission event. The simian immunodeficiency virus, found in chimpanzees, is believed to be the precursor to HIV, highlighting the complex evolutionary dynamics of viral pathogens and their impact on human health (MedicalNewsToday, n.d.).

2.3 Retrovirus vs. Lentivirus (HIV) Virion Components

HIV is a member of the *Retroviridae* family, categorized explicitly as a lentivirus, and is notorious for its ability to cause acquired immunodeficiency syndrome (AIDS) if left untreated. The pathogenicity of HIV arises from its distinctive process of reverse transcription, wherein the viral RNA genome is enzymatically converted into DNA by the action of reverse transcriptase. Subsequently, this newly synthesized DNA moiety infiltrates the host cell nucleus, where it integrates into the cellular genome, assuming control and adopting the status of a provirus. Under the regulation of the host's DNA-dependent RNA polymerase II, transcription of the proviral DNA occurs, yielding multiple RNA copies. These RNA transcripts serve as templates for two distinct processes: one culminating in the assembly of progeny virus genomes, facilitating viral replication. At the same time, the other directs the synthesis of messenger RNAs (mRNAs). This orchestrated series of molecular events ultimately leads to the generation of viral progeny, perpetuating the infection cycle and contributing to the progressive depletion of host immune defenses characteristic of HIV pathogenesis. (Dulbecco, 1988). Delving into the comparative genomics of retroviruses,

explicitly contrasting the retroviral genome with that of HIV (a type of lentivirus), provides a deeper understanding of viral diversity and pathogenesis within the *Retroviridae* family. Retroviruses, including HIV, share common genomic characteristics characterized by single-stranded RNA genomes. However, notable differences exist between the general retroviral genome and that of HIV, particularly in genome organization and key genetic elements. The general retroviral genome typically comprises three essential genes: gag, pol, and env, encoding structural proteins, enzymes essential for replication, and viral envelope proteins, respectively. Additionally, regulatory elements such as the long terminal repeats (LTRs) flank the viral genome, serving as promoters for transcription and integration sites during the viral life cycle (A D Frankel, 1998). In comparison with HIV, a member of a lentivirus has a distinct genome that reflects its pathogenicity and replication strategies. HIV has gag, pol, and env genes, but also includes regulatory and auxiliary genes including tat, art, and sor that play important roles in viral replication, immune evasion, and pathogenesis. The HIV genome has complex regulatory sections, including long terminal repeat (LTR), which comprises sequences necessary for transcriptional regulation, viral integration, and latency establishment (R Carroll, 1991).Comprehending and learning the differences between the general retroviral genome and HIV's original genome can provide great insights in the scientific field on the molecular mechanisms at work. HIV pathophysiology, viral replication dynamics, and prospective treatment targets. Identifying genetic differences can help researchers understand the problems of HIV infection and design more effective prevention, treatment, and cure techniques. In lentiviruses, viral RNA is a single-stranded positive-sense molecule measuring 10 kb and varying in size.

- Viral RNA is supported in both structural integrity and functional efficiency within host cells by its 5' cap and 3' poly-A tail. Just as important are the genes that are embedded in between these caps they code for proteins that are essential at different points in the viral life cycle. Several other sequences, including the dimerization linkage structure and the primer binding site, are essential for controlling how the virus replicates and spreads throughout host cells.
- The gag, pol, and env genes are usually found in retroviruses, and they are arranged from the 5' cap to the 3' poly-A tail. This is known as gene organization, and it is essential for

producing the critical proteins required for the virus's assembly, replication, and penetration into host cells.

- The terminal redundancy (R) region of the retroviral RNA, which is essential for the start of reverse transcription and efficient genome packaging, comes next to the cap structure at the 5' end of the RNA. The primer binding site (PBS) and the U5 sequence, which come before this region, are necessary for reverse transcription and the integration of the viral genome into the host's DNA.
- Two RNA molecules are joined by a dimerization linkage structure (DLS), which guarantees their inclusion in the virions created during assembly. The genetic stability and efficient spread of the virus depend on this relationship and virus integration processes, enabling proviral DNA transcription, integration into the host genome, and the start of reverse transcription (R Swanstrom, 1997).

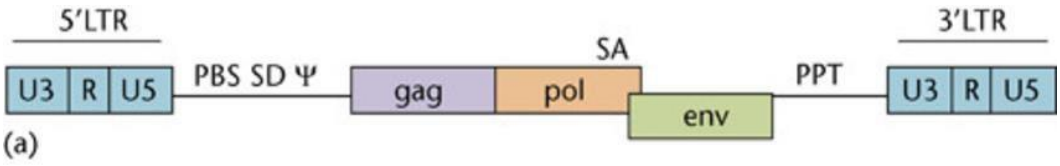

Figure 2: Elements of Retroviral genomic architecture (Wilber, 2014).

2.4 Contrasting the Genomic Features of HIV

HIV contains more genes than the typical retroviral genes gag, pol, and env. These include: ORF (open reading frame), sor, 3' ORF, Tat, Art, and These genes play distinct roles in HIV replication and pathogenesis Sor participates in regulating viral replication. Art functions as an anti-repressor trans activator, boosting translation of the env and gag genes. Furthermore, the 3' ORF regulates numerous events in the viral life cycle, hence limiting viral replication. These genes are essential for HIV replication, viral gene expression, and immune evasion (Yasemin van Heuvel, 2022).

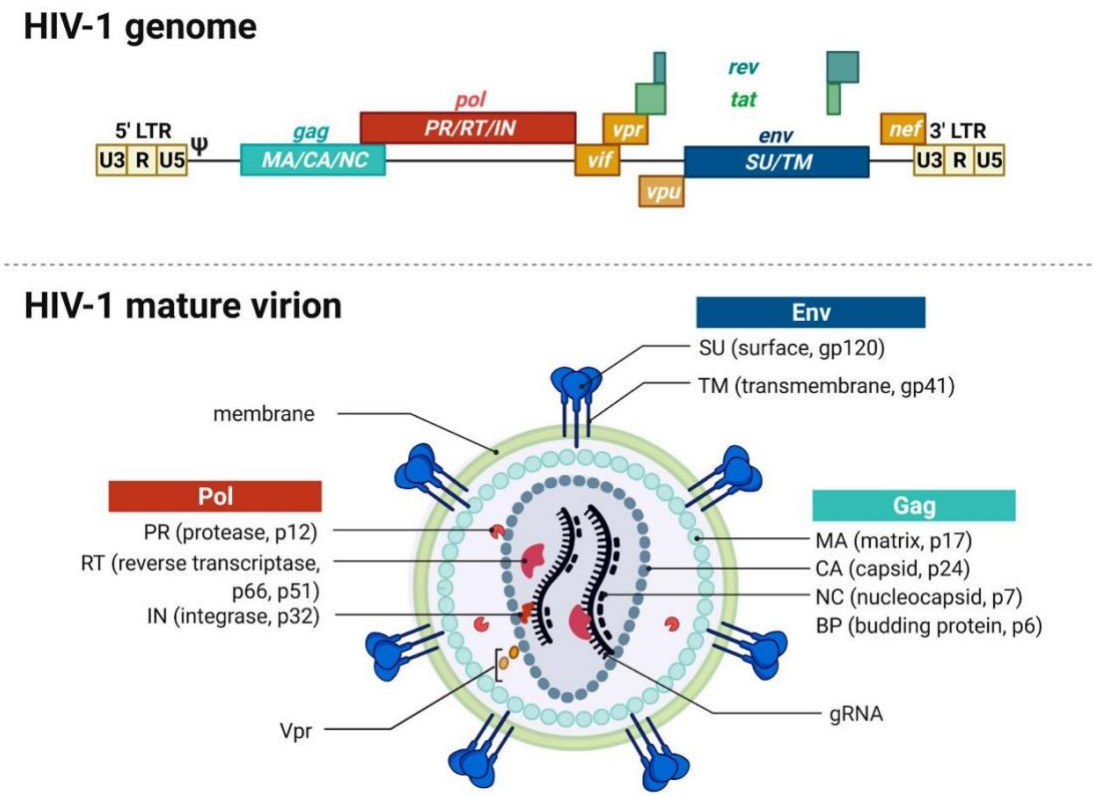

Figure 3: Genome composition of HIV-1 and virion morphology Insights (van Heuvel, 2022).

2.5 Discovery of VDR Gene

Vitamin D's discovery in 1920 highlighted its crucial role in maintaining calcium levels in bones and the intestine. By 1932, its chemical structure revealed its steroid nature. However, a groundbreaking revelation in the late 1960s identified it as a precursor to a novel steroid hormone, [1α,25(OH)$_2$D$_3$], synthesized by the kidney acting as an endocrine gland. This discovery, in 1969, of the Vitamin D receptor (VDR) for [1α,25(OH)$_2$D$_3$] marked a transformative moment in science. (Norman, 2006).

Over the following two decades, research proliferated, elucidating the presence of VDR in over 30 human organs and tissues, driven by the influence of the vitamin D endocrine system. Advancements in genomics further revealed the cellular distribution of VDR, encompassing the immune system, hair follicles, bone marrow, lymphocytes, adipose tissue, and cancer cells. Still, there was a dilemma with some physiological processes, such as the quick migration of endothelial cells, the opening of calcium and chloride channels in osteoblasts, and the quick absorption of

calcium in the intestine. These rapid reactions raised questions about whether a second receptor—different from the nuclear VDR—mediated these quick responses. The quest for solutions produced some fascinating discoveries. Although VDR mostly dwells in the nucleus of cells, in specific cells, it is also linked to caveolae in the plasma membrane. Furthermore, [1α,25(OH)$_2$D$_3$]'sconformational flexibility enables it to provide a variety of ligands for VDR, specifically promotingfast or genomic responses. (Norman, 2006). An ensembl model on Error! Reference source not f ound. of VDR clarified how diverse ligands, such as [1α,25(OH)$_2$D$_3$], operating through VDR at distinct cellular sites, coordinate rapid and genomic responses. Our knowledge of the complex rolesthat vitamin D plays in cellular physiology and signaling pathways has increased as a result of thesediscoveries.

2.6 Structural Features of 1α,25(OH)$_2$D$_3$: Conformational Flexibility

A. The side chain of 1α,25(OH)$_2$D$_3$ is confirmation of the flexibility, which is shown by the existence of five single carbon-carbon bonds that are depicted by the five arrows.

B. 1α,25(OH)$_2$D$_3$'s cyclohexane-like A-ring and its quick switch between two chair-chair conformers. Millions of times per second, this exchange leads to the 1α- and 3-hydroxyls equilibrium between axial and equatorial orientations. Due to this dynamic movement, the hydroxyl groups (1α and 3) can shift between locations, pointing either inward or outward, allowing the molecule to adopt alternative orientations.

C. 1α,25(OH)$_2$D$_3$ can adopt several conformations due to its rotational freedom around the 6–7 carbon-carbon single bond of the seco B-ring. Among these is the 6-s-trans conformation, which is open and stretched and preferred for genomic reactions mediated by the vitamin D receptor (VDR). Furthermore, a more compact conformation known as 6-s-cis, which has a steroid-like structure, is known to be used by the VDR for rapid reactions.

D. The structure of JN, which has the same effectiveness as 1,25(OH)2D$_3$ in inducing rapid reactions and is chemically bound in a permanent 6-s-cis form.

E. Space-filling (E) and stick (F) represent the apparent differences in the shapes of the three optimal ligands for the nuclear-localized VDR (left), the membrane-caveolae localized VDR (middle), and the plasma DBP (right) (Norman, 2006).

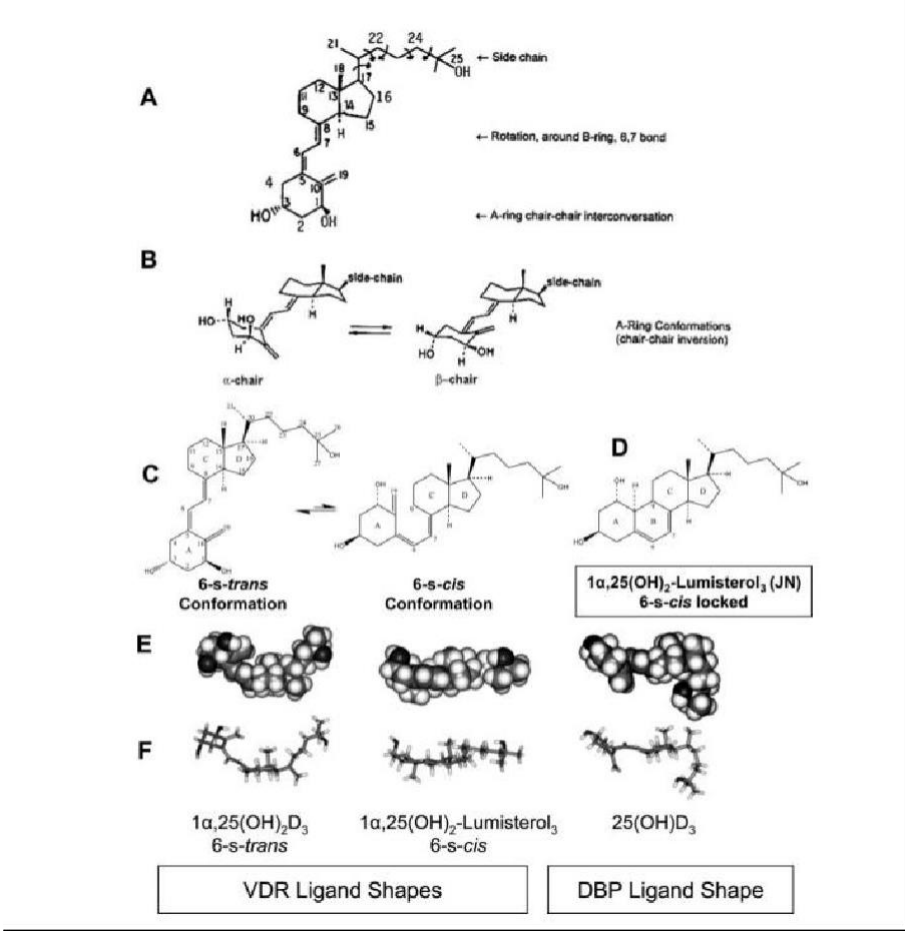

Figure 4: Dynamic conformations and functional consequences associated with 1α,25(OH)$_2$D$_3$ and its analogues. (Norman, 2006)

2.7 VDR Structure: Ligand Binding & Cellular Responses

In the mid-1980s, researchers discovered 1α,25(OH)$_2$D$_3$-mediated rapid responses (RR), which were previously unknown. Such rapid responses raised more questions, because genomic responses of the VDR took way longer, and it was odd to find rapid responses. These rapid responses, which occurred within minutes to an hour, defied explanation as simply the result of nuclear Vitamin D receptors (VDRs) regulating gene transcription. Rapid calcium absorption in the gut (transcaltachia), insulin release by pancreatic beta-cells, activation of voltage-gated calcium and chloride channels in osteoblasts, and rapid endothelial cell migration are all examples of rapid responses.

Following initial setbacks, data emerged indicating that the conventional VDR, which is traditionally associated with the cell nucleus, can also be found in caveolae on the plasmamembrane in some cells. Furthermore, the conformational flexibility of $1\alpha,25(OH)_2D_3$ allows it to create ligands of varied forms for the VDR, selectively targeting either genomic or RR pathways. This mini view proposes a proposed conformational ensemble model of the VDR. It sheds light onhow varied ligand forms of $1\alpha,25(OH)_2D_3$, acting through the VDR at multiple cellular sites, can efficiently control both genomic and RR responses. This model eventually gives us an answer on how rapid responses occur.

Error! Reference source not found. Below is the suggested conformational ensemble model of the V DR which provides insights into how varied ligand shapes and receptor conformations caninfluence cellular responses and RR.

A. The primary focus lies on the ligand binding domain (LBD) of nuclear VDR. X: The LBD structures, depicted as ribbon structures, have shown differences in their interior shape and volume in comparison to the structure of $1\alpha,25(OH)_2D_3$. Y; is an illustration for the VDR showing the locations of a putative alternate pocket (blue: 1,25(OH)2-lumisterol, a nongenomic 6-s-cis locked agonist) and the classic genomic pocket (red: 1,25(OH)2-vitamin D_3, a genomic agonist) as determined by molecular modeling. Two ligand-binding pockets are shown: the classic genomic pocket and a putative alternative pocket. While the alternative genomic pocket could be used for nongenomic agonists like 1,25(OH)2-lumisterol, the standard genomic pocket can hold genomic agonists such as $1,25(OH)2D_3$. The closed conformation is represented by the brown helix 12 (H12). The alternate pocket portal that has been suggested is located between the H1 C terminus and the H3 N terminus. It is interesting to note that both pockets stabilize their 1- and 3-hydroxyls by using the same hydrogen-bonding partners (R274, S237, and S278), indicating that the VDR is unable to bind two ligands at once (Norman, 2006).

B. In this section, a receptor ensemble schematic model is presented, implying the existence of three distinct unoccupied receptor species (a, b, and c) in rapid equilibrium with one another. Distinct places for these receptor conformer species represent Helix 12 (H12). Differently shaped ligands may be preferentially bound by each conformer. A ligand may initiate rapid

reactions (RR) if it occupies the alternative pocket; on the other hand, it may activate genomic responses if it occupies the classical pocket. All three H12 conformers allow access to the alternate pocket, which is preferred for occupancies before reaching steady-state equilibrium, particularly for the receptor's endogenous ligands. Only conformers (c) can receive ligands that bind to the standard ligand binding site and trigger genetic reactions. The nuclear receptor's attachment to caveolae or cell membranes can shift the preferences of ligand binding, favoring the other pocket. The effectiveness of cellular signaling pathways mediated by either pocket may be impacted by drug analogs, especially those that are connected to natural hormones, as they may have differing fractional occupancies of the two pockets (Norman, 2006).

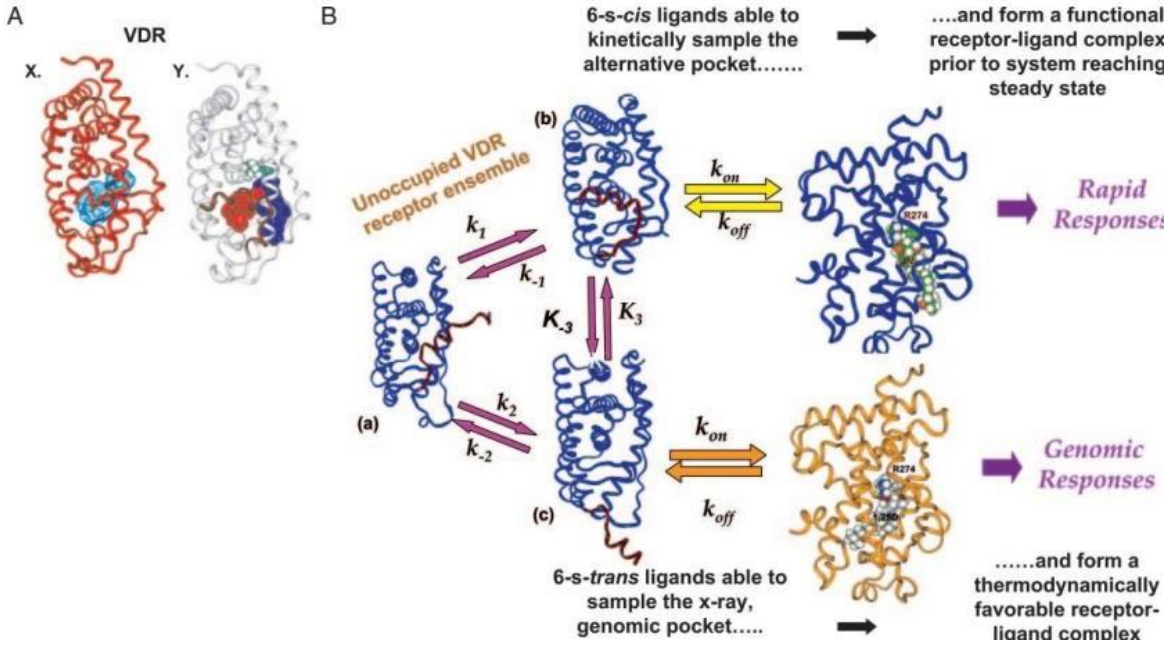

Figure 5: Structural details of the VDR that support a conformational ensemble model.

2.8 VDR Gene Overview

Vitamin D receptor (VDR) is a member of the nuclear receptor family, specifically the steroid hormone receptor superfamily. In the genomic context, its location is 12q13.11, with an exon count of 12. These receptors are activated by particular molecules (ligands) and regulate the expression

of specific genes. In the case of VDR, it binds to a hormone called $1\alpha,25(OH)_2D_3$, which is the active form of vitamin D. This binding initiates a series of events that control the activity of genes involved in vital biological processes such as bone metabolism, phosphorus balance, and immune function. The primary role of VDR, along with $1\alpha,25(OH)_2D_3$ and vitamin D in general, is to maintain the balance of minerals, particularly calcium, in the body. Issues with the Vitamin D Receptor (VDR) can lead to problems such as hypocalcemia and weakened bones due to insufficient mineralization. Vitamin D_3, known also as cholecalciferol, is produced by the skin following sunlight exposure and is important for maintaining healthy bones. The hormone 1α,25-dihydroxyvitamin D_3, an active form of Vitamin D_3, is key in facilitating the absorption of phosphorus and dietary calcium. The interaction between VDR and $1\alpha,25(OH)_2D_3$ starts a sequence of molecular actions that affect genes important for a range of bodily functions, thus supporting bone health and immune function, among other essential processes (MacDonald, 2004).

The mechanism of action includes the following steps of Error! Reference source not found..

1. $1\alpha,25(OH)_2D_3$(calcitriol) and 25-$(OH)D_3$ (calcidiol) are bound to a specific carrier protein known as vitamin D binding protein (DBP) while circulating in the bloodstream.
2. The conversion of 25-(OH) D_3 to $1\alpha,25(OH)_2D_3$ takes place in the cytoplasm, which is mediated by an enzyme known as CYP27B1- leading to the activation of VDR.
3. Once it has been produced, $1\alpha,25(OH)_2D_3$ binds to the Vitamin D Receptor (VDR) protein. VDR acts as a transcription factor, regulating the expression of specific genes in response to binding with 1,25-(OH)2D_3.
4. VDR-$1\alpha,25(OH)_2D_3$ complex heterodimer with RXR (Retinoid X Receptor)
5. VDR/RXR complex binds with VDRE (located in promoter regions of selected genes).
6. VDR/RXR recruit SRCs and p300/CBP as co-activators for their HAT (histone acetyltransferase) activity loosens the chromatin structure.
7. SRCs and p300/CBP HAT (histone acetyltransferase) activity trigger recruitment of other co-activators such as NCoA62/SKIP and Mediator-D multimeric complex (depicted as co-modulators in the picture).
8. Mediator-D multimeric complex brings in RNA polymerase II and the core transcriptional machinery to the target gene's promoter region, commencing active transcription of the gene (MacDonald, 2004).

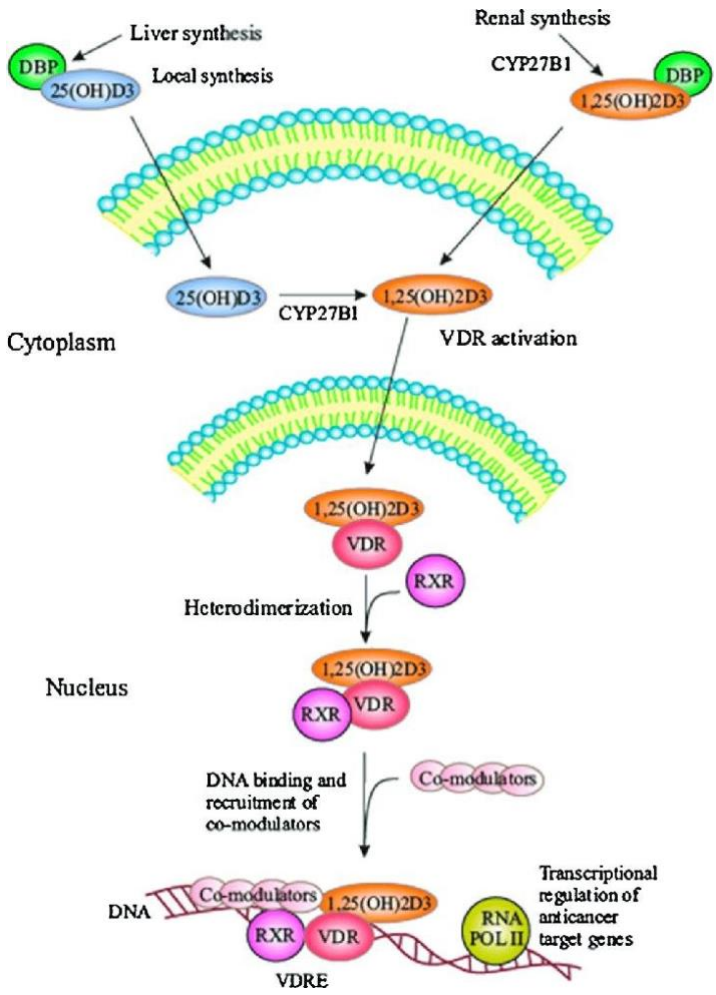

Figure 6: VDR-related genomic mechanism of calcitriol action. (Fathi, 2019)

Therefore, when VDR binds to 1α,25(OH)$_2$D$_3$, it sets in motion a series of molecular events that influence genes crucial for various physiological functions, ensuring proper bone health and immune response, among other vital processes.

2.9 Studies on VDR Variants Connection with HIV Progression

Recent studies have shed light on the relationship between Vitamin D$_3$, its active form 1α,25(OH)$_2$D$_3$, and a cellular process known as autophagy, which involves the degradation and recycling of cellular components. These studies have revealed that HIV-1, the virus responsible for

AIDS, reduces autophagy during its replication cycle. This reduction in autophagy compromises the ability of cells to eliminate viruses can affect how quickly HIV infection spreads. Significantly, studies suggest that vitamin D_3 in its active form may increase autophagy and decrease HIV-1 replication (Stephen A. Spector, 2011). Examining the course of the illness in people with low vitamin D levels or mutations in the vitamin D receptor (VDR) gene that reduce vitamin D's effectiveness could provide a novel HIV diagnostic strategy. The functions of VDR, Vitamin D_3, and its active derivative, $1\alpha,25(OH)_2D_3$, are interdependent, indicating that thorough investigations into their separate and combined roles may enhance our comprehension of HIV's origins. Clinicians can gain important insights into the dynamics of HIV infection and treatment by how it impacts the immune system. Improved diagnostic instruments, improved tracking of HIV progression, and identification of high-risk individuals could all result from this kind of research. Moreover, this information may help develop specific treatment plans that increase VDR activity and raise vitamin D levels to strengthen HIV-specific immune responses.

2.10 Genetic Variants in VDR Gene and HIV Progression

A 2004 case-control study examines the relationship between the folk-1 polymorphism in the VDR gene and the role of HIV seropositive patients' development to AIDS. In this study, 185 HIV-1-positive white Caucasian patients who were at risk of intravenous drug consumption were evaluated. The Centers for Disease Control (1993) criteria were followed by the researchers in their primary assessment of the course of aids, with the first declines in CD4 cell counts being as low as 200 cells/μL. Those patients who achieved these results were labeled as progressors, whereas those who did not achieve any outcomes after at least eight years of follow-up were labeled as non-progressors. According to the study, heterozygous individuals (Ff) for the Fok-I polymorphism were more prevalent in the groups who developed AIDS and saw a decrease in CD4 cell count to less than 200 cells/Additionally, compared to people with other genotypes, those with the Ff genotype demonstrated a shorter time to assistance. Furthermore, compared to non-Ff patients (such as ff or FF), the decline in CD4 cell count to less than 200 cells/μL occurred more quickly in Ff carriers. These findings lead the study to hypothesize that individuals with the Ff genotype of the Fok-I polymorphism who are HIV-1 seropositive may be more susceptible to developing AIDS rapidly. As a result, this study uses a case-control methodology to investigate the connection between an HIV-1 patient's genetic polymorphism and the progression of their infection. While this

case study offered valuable insights, its focus on using exclusively white Caucasian patients recruited between 1982 and 1991 limits its generalizability to broader populations. It includes various age groups for better understanding. (Gema Nieto, 2004). In another recent case study that took place in 2013, researchers delved into the association between variants of the Vitamin D Receptor (VDR) gene and HIV infection in children. They focused on specific variations, such as Fok-1 and Bsm-1, known for their impact on HIV progression in adults. These variations affect how VDR functions at a molecular level. The study looked at five different VDR variants, including rs2228570 (Fok-I C/T) and rs1544410 (Bsm-I G/A), which are known to influence VDR activity. Additionally, three more SNPs were also examined: rs12785878 (DHCR7/ NADSYN1 G/T), vitamin D transport: rs2282679 (GC A/C), and vitamin D hydroxylation: rs10741657 (CYP2R1 G/A).

Results revealed that specific variants, particularly those associated with Fok-1 and Bsm-1, were linked to HIV disease progression in children over the age of two. Interestingly, this association was more pronounced in Hispanic and white children under two years old. However, rs2282679 (GC A/C) and CYP2R1 G/A did not show a significant association with HIV progression. This highlights the complexity of genetic factors involved in HIV infection. Overall, the study provides valuable insights into how specific variations in the VDR gene can influence HIV disease progression in children. By including children of different races and ages, the researchers offer a comprehensive understanding of these genetic associations, with one major limitation being that nutritional intake and consumption of vitamin D levels were not available, meaning this study could not assess the vitamin D levels in association with HIV disease progression (Amaran Moodley, 2013).

2.11 Exploring rs746619116 as a Diagnostic Marker for HIV+ in Pakistan.

In response to the escalating HIV prevalence in Pakistan, our study addresses a crucial research gap regarding the role of VDR gene variants in HIV disease progression within this demographic. Our primary objective is to assess the potential of a specific VDR variant, rs746619116, as a diagnostic marker for monitoring HIV progression in the Pakistani population. This variation, which replaces arginine with tryptophan at codon 154, has the potential to accurately predict disease progression from what our *in silico* results indicated. Our research on the Pakistani population aims to provide

valuable insights for developing personalized HIV diagnostic techniques. We want to develop early diagnosis and intervention options for VDR variant rs746619116, leading to better results, which can be of benefit to Pakistani society.

Chapter 3: Methodology

3.1 *In silico* Tools

The protein sequence of VDR, downloaded from the ENSEMBL (Fergal J Martin, 2023) database in FASTA format corresponding with the transcript ID: ENST00000395324.6 compasses 427 amino acids. Employing AlphaFold, the amino acid sequence of VDR was used to construct its structural model followed by domain predictions inspected through InterPro. Pathogenic variants for VDR were obtained from ENSEMBL. From this selection, 28458 missense variants were selected based on the connection which disease then the pathogenicity percentage of each missense variant was determined by using 6 common tools including REVEL (Nilah M. Ioannidis, 2016), Mutation Assessor (Boris Reva, 2011), metaLR, Sift (Pauline C Ng, 2003), Polyphen2 (Ivan Adzhubei, 2013), and CADD (Martin Kircher, 2014) to give a pathogenicity score. Then variants with pathogenicity crossing 70% were considered after strict and detailed scoring standards were implemented by multiple common tools which resulted in 3 variants. Next SNP with the greatest pathogenicity score was isolated for further investigation.

3.2 Structure Retrieval and *In silico* Mutagenesis

PyMOL version 2.5.8 (Schrödinger, 2024) was employed to display a visual representation of the SNP's influence on the 3D structure of the protein. The mutation was introduced in the structure by placing the amino acid sequence for the mutant in the original VDR wild-type structure. By doing this we were able to see and compare the wildtype and mutant 3D structure side by side allowing us to point out how the substitution impacted the configuration of the protein. Which could impact the functional role of the protein.

3.3 Stability Study of SNP

The impact of this variant on the stability of the protein structure was assessed using I-Mutant 2.0 (Emidio Capriotti, 2005)The impact of SNP on protein stability was assessed using the I-Mutant software, which provided delta delta G values. An increase in stability correlates with a higher delta G (ΔΔG) value, while a decrease indicates an energy reduction, potentially affecting normal protein pathways. The results of SNP-induced changes in stability were predicted based on these (ΔΔG), s.

Maestro (Laimer J, 2015 Apr 16)was also used to examine the impact of a specific SNP on the VDR protein, allowing for an analysis of changes in protein shape and stability post-mutation. This software provided data on the free energy change (ΔΔG).

3.4 Structural and Functional Study of SNP

Project HOPE (Have (y)Our Protein Explained) was utilized in our studies as a bioinformatics tool designed to analyze and illustrate protein functions and structures, primarily through amino acid sequences or 3D structural data. This tool proved extremely valuable for visualizing the effects of specific amino acid mutations. It accomplished this by displaying the original amino acid's schematic structure and its transformation into the mutant form, highlighting changes such as size, hydrophobicity, and the potential loss of hydrogen bonds or salt bridges due to intra- or inter-residue interactions. Project HOPE pinpointed the location of a mutation within a protein domain and linked it to a specific variant through the use of dbNSFP. Additionally, it included a metaRNN score to assess the possible pathogenicity of the variant (Chang Li, 2022). This tool was utilized to increase the understanding of how mutations impact protein function and their role in disease progression. Moreover, another software, DynaMut 2 (Carlos H.M. Rodrigues, 2020), was used to find out the effects of alterations in a protein's amino acid sequence on its stability. It offered detailed information on how these changes can alter the protein's structure and integrity. The software provided multiple calculations, such as changes in free energy (ΔΔG), which helped determine whether a mutation decreased protein stability (indicated by positive ΔΔG values) or increased it (reflected by negative ΔΔG values).

3.5 VDR gene Variant rs746619116 Association with HIV

The rs746619116 variant in the VDR gene was analyzed to determine its relationship with HIV infection using the software known as FATHMM, which makes use of statistical models such as Hidden Markov Models (HMMs) (Hashem A Shihab, 2013; Rogers MF, 2018). With the ability to predict variants, this method is intended to assess and project the impact of genetic variants on protein functionality. By using this software, we intended to understand whether the rs746619116 variant, which is concluded to be pathogenic, has any link to HIV. This tool is mainly used in genetic research to investigate how mutations might impact protein functions and to examine their potential

health implications (Hashem A Shihab, 2013). Another tool that was used to predict if the structure changes due to a mutation had any association with the disease was MutPred2 (Vikas Pejaver, 2020) this tool incorporates both molecular and genetic information to find out how the mutation might affect the protein function. It examines changes in structural components and how mutation may affect molecular interactions. It provides a numerical score that represents the likelihood of the mutation causing disease with a higher score demonstrating higher risk.

3.6 Cellular Localization

Deeploc 2.0 (Vineet Thumuluri J. J., 2022) software's main role is in protein cellular localization estimation. It uses protein sequence examination to identify locations for proteins within a cell, such as the cytoplasm, cell membrane, or other cellular compartments.

3.7 Primer Designing

The primers were designed to identify the specific genetic variant present in the samples and for that purpose an online web tool Primer1 (Andrew Collins, 2012) was employed and the primers for ARMS PCR were designed. To obtain the ARMS PCR primers using Primer1, the genome sequence input was provided as shown in Error! Reference source not found.**.** ARMS PCR primers obtained from Primer1. The location of the VDR on the human chromosome was identified as 12:47857506. For the tool, alterations were made to the allele difference and SNP position, while keeping the remaining settings the same.

Table 1: ARMS PCR primers obtained from Primer1.

Name	**Sequence**
Forward Inner Primer	CCGACTTCTGCCAGTGCT
Reverse Inner Primer	AGCAGGCAGACATACACG
Forward Outer Primer	TCTGACAGATGAGGAAGTGCA
Reverse Outer Primer	GAAGGGAAAGATGGGGTCTC

3.8 Patient Inclusion and Exclusion Criteria

For the genetic studies, 50 control patients without any diseases were included, and their ages were recorded this included individual above and below the age of 40. Additionally, 50 HIV-positive patients were also selected and data on their average age, vitamin D levels, and platelet counts were collected. Transgender and thalassemia patients that were undergoing blood transfusion and individuals that had other disease as well like herpes or hepatitis C were not excluded.

3.9 DNA Extraction

Began with the collection of 50 control blood samples from individuals without HIV. Subsequently, DNA extraction was performed over a two-day period. On Day 1, each blood sample (750 µl) was mixed with an equal volume of Solution A (1:1 ratio) in an Eppendorf tube and allowed to incubate at room temperature for approximately 10 minutes. After centrifugation at 13000 rpm for 10 minutes, the supernatant was discarded, and Solution A (400 µl) was added again. Following another centrifugation step at 13000 rpm for 10 minutes, the supernatant was discarded, and Solution B (430 µl) was added. Mechanical mixing was employed until the pellet disappeared, and the mixture was centrifuged again at 13000 rpm for 10 minutes. Then, 12 µl of SDS and 5 µl of Proteinase K were added, mixed well, and the tubes were placed in an incubator overnight at 37°C.On the second day solution C (250 µl) and Solution D (250 µl) were added to the incubated samples, followed by centrifugation at 13000 rpm for 10 minutes. The aqueous layer containing DNA was transferred to a separate tube. DNA precipitation was induced by adding 55 µl of 3M Sodium Acetate and 500 µl of chilled isopropanol, followed by centrifugation at 13000 rpm for 10 minutes. The DNA pellet obtained after centrifugation was resuspended in 200 µl of chilled ethanol and then centrifuged at 13000 rpm for 8 minutes to remove excess ethanol. After air drying to evaporate residual ethanol, the DNA was suspended in 250 µl of PCR Water for further analysis.

3.10 ARMS PCR

After DNA extraction, the study proceeded to conduct Allele-Specific PCR Amplification Refractory Mutation System Polymerase Chain Reaction. One technique used to specifically amplify DNA segments containing a specific allele variant is called ARMS-PCR. Each blood sample used in this procedure had a PCR tube set up with the proper labelling for identification. Two PCR tubes were assigned to each sample: one labelled C for the wild-type allele and another labelled T for the mutant allele. For this setup, two PCR tubes had to be made for every 50 samples. Each tube was filled with 1 µl of a common primer, 3 µl of PCR water, 6 µl of PCR master mix, and 1 µl of the DNA sample, totaling 11 µl. To differentiate the tubes, 1 µl of the wild-type allele-specific primer (C) was added to one tube, and 1 µl of the mutant allele-specific primer (T) was added to the other, bringing each tube's final volume to 12 µl. This preparation was consistently carried out across all samples.

3.10 ARMS PCR Conditions

Initially, the PCR mixture was heated to 94°C to facilitate DNA denaturation. Subsequently, the temperature was lowered to 50–60°C, allowing for the rejoining of single DNA strands and primer attachment. DNA synthesis commenced as the temperature was raised to 72°C, optimal for Taq polymerase activity. During this initial PCR stage, long products were generated from each DNA strand, possessing identical 5′ ends but random 3′ ends due to chance termination of DNA synthesis. In subsequent cycles, short products were formed, with both 5′ and 3′ ends determined by primer annealing positions. Final elongation step was performed at 70–74°C for 5–15 minutes after the last PCR cycle to ensure complete elongation of any remaining single-stranded DNA (K Mullis, 1986)

3.11 Gel Electrophoresis

After the PCR amplification, gel electrophoresis was performed to visualize the amplified DNA fragments. Gel electrophoresis is a technique used to separate and analyze DNA fragments based on their size and charge. 2 grams of agarose powder were measured and added to a beaker filled with 100 mL of TAE buffer. This mixture was heated in a microwave until it became transparent. Next, 7 µl of ethidium bromide, a carcinogenic substance, was carefully added to the clear solution, with the precaution of wearing gloves and working in an isolated area. The solution was then gently

stirred. While the solution was cooling, combs were placed into the casting tray to create wells, and the agarose solution was poured into the tray to set. The gel was left to solidify for 20 minutes before being placed in a buffer. Subsequently, 6 µl of the PCR product from each tube was loaded into individual wells on the gel. The electrophoresis apparatus was set to run at 80 volts and 500 mA for 20 minutes. After the run, the gel was placed under a UV transilluminator to visualize the DNA bands, allowing for the observation of the separated DNA fragments.

Chapter 4: Results

4.1 Prediction of the Pathogenic VDR Variants

To ensure the precision of variant analysis for the Vitamin D receptor (VDR) gene, a collaborative method was utilized. Initially, the ENSEMBL database was employed to collect all known variants linked with the VDR gene. Following this, variants from two distinct databases, namely COSMIC and gnomAD, were integrated to supplement the dataset.

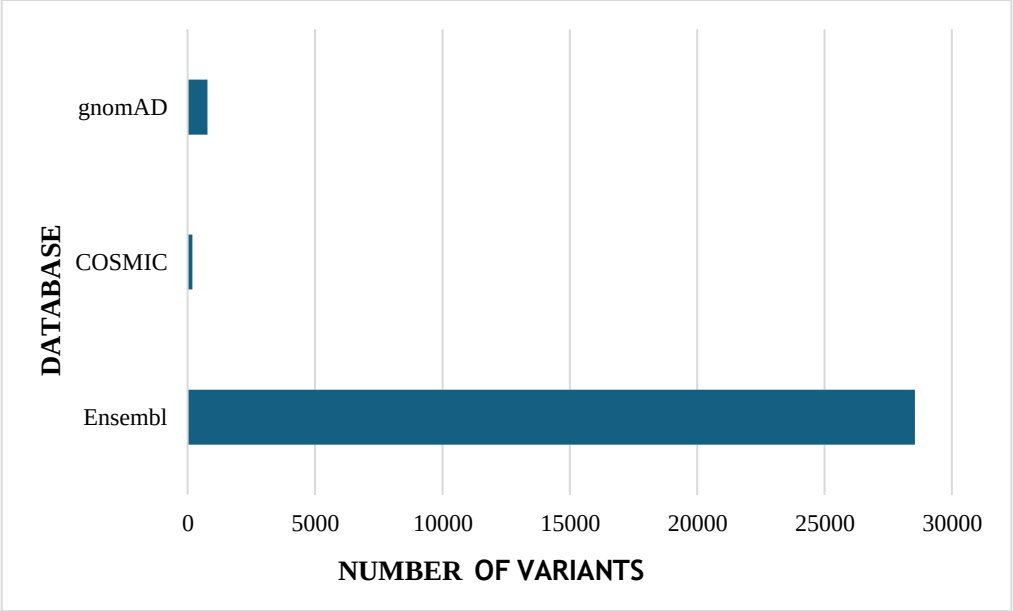

Figure 7*:* Summary of the variants retrieved from Ensembl, COSMIC, and gnomAD

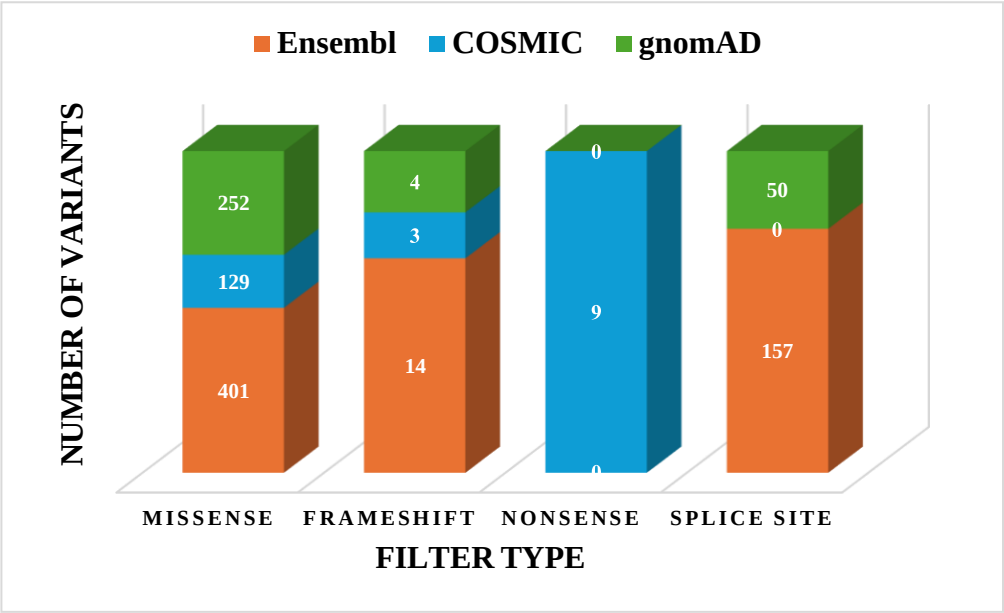

Figure 8: Number of variants from different databases following filtration.

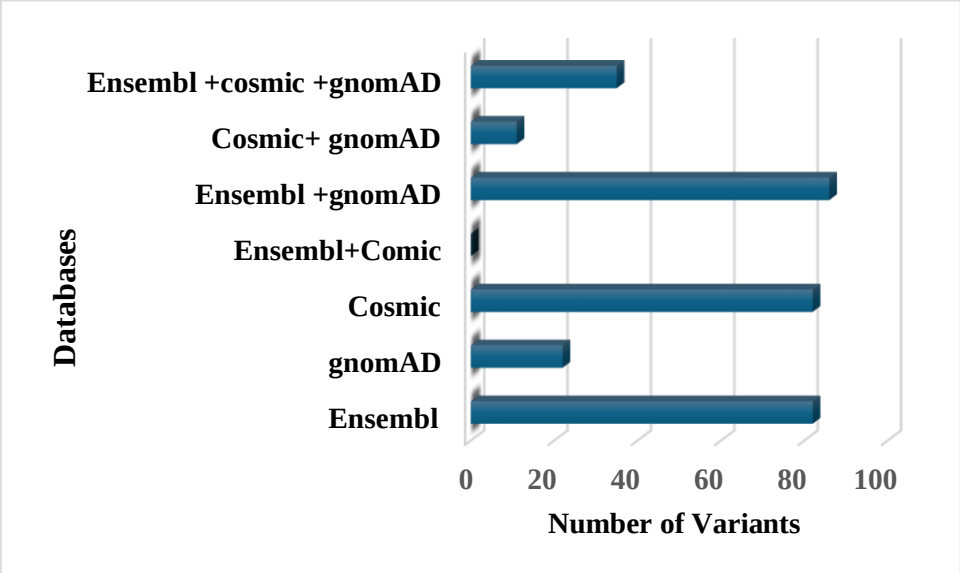

Figure 9: Number of Common and unique missense variants of the VDR gene obtained from different databases.

Figure 7 shows a comprehensive summary of the variants retrieved from the ensemble database, COSMIC, and gnomAD, outlining the variants exhibited by each database. This collaborative approach, alongside the incorporation of multiple databases, guaranteed a more comprehensive and complete dataset for precise analysis of VDR gene variants. Following the retrieval of variants from different databases, several filters were applied to them, beginning with missense mutations, frameshift, nonsense, and lastly, splice-site mutations. **Figure 8** shows the number of variants that remained in each database after these filters were added. Next, the common and unique variants among the databases were identified figure shows the results. **Figure 9** shows the unique and common VDR missense single nucleotide polymorphism from ensemble gnomAD and comic databases. Ensembl and cosmic had the highest unique missense SNPs gnomAD had the least. Similarly, the ensemble and gnomAD had the most common SNPs. The Ensembl genome browser was employed to identify 401 missense variants, each with its rsID. The missense variations were examined to determine the percentage of SNP effects to do this their pathogenicity was determined, and six consensus tool classes were utilized These six tools were SIFT , PolyPhen, CADD, Revel, and MetaLR these tool classes classified the variants as deleterious or tolerant the deleterious variants were given a score of 1 while the tolerant variant was given a score of 0 these scores were added up and divided by the number of tools used and then multiplied by 100 to calculate the pathogenicity percentage of each variant and pathogenicity graph containing the pathogenicity percentage was constructed **Figure 10** shows the pathogenicity graph.

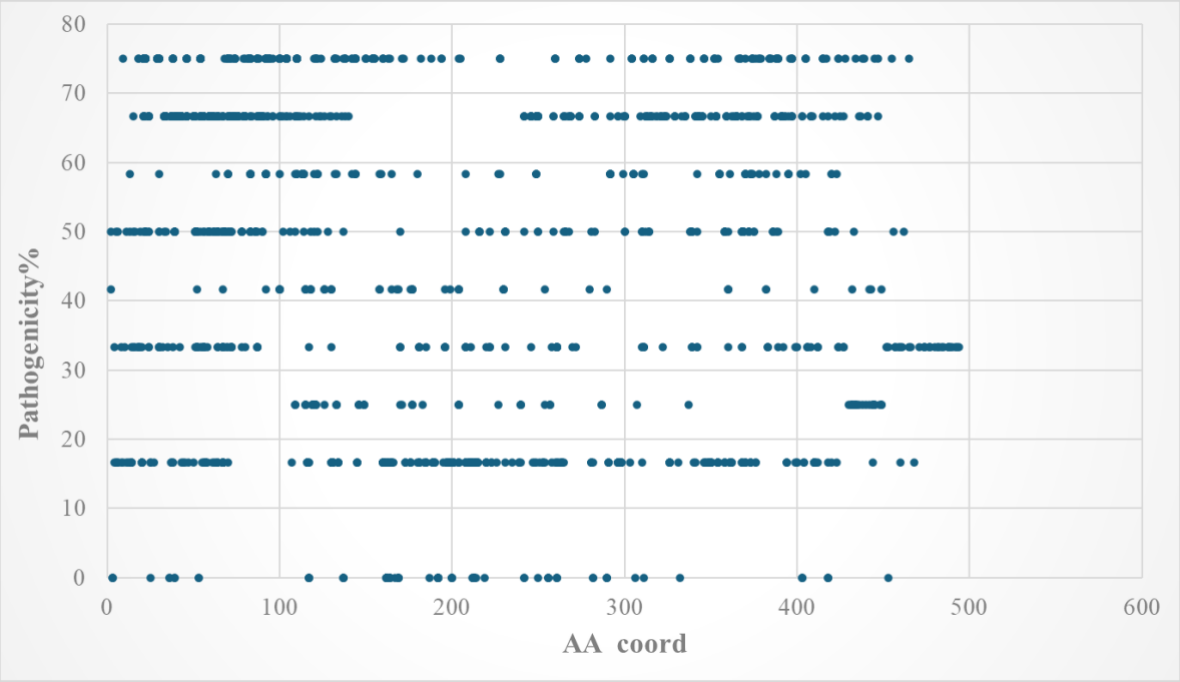

Figure 10: Pathogenicity percentage of missense variant.

Following this, the missense variants that had pathogenicity over 70% were selected and tools scoring was applied to them. This resulted in 3 variants that had the highest pathogenicity score determined by tools Detailed information about these variants, including their rsIDs, chromosomal locations, alleles, modified amino acids, and positions of these amino acids, is presented in **Table 2.** Then among these 3 variants, the one with the highest pathogenicity score determined by multiple common tools score was selected which was rs746619116 shown in **Table 3.** These six tools were assigned based on the pathogenic score. SIFT (Sorting Intolerant from Tolerant) predicts whether amino acid substitutions are harmful or not. Scores below 0.1 suggest that the substitution is deleterious, which means it negatively affects protein function. The second tool PolyPhen (Polymorphism Phenotyping PolyPhen) determines the impact of amino acid changes on protein structure and normal function. Scores above 0.95 classify the variant as probably damaging. CADD (Combined Annotation Dependent Depletion) CADD ranks variants according to their predicted deleteriousness. A score of 30 or higher indicates a likely harmful variant. Revel (Rare Exome Variant Ensemble Learner). Revel is an ensemble-based predictor that highlights potentially damaging rare missense variants. Scores above 0.5 suggest damaging effects MetaLR (Meta Logistic Regression) MetaLR merges multiple scores through logistic regression to predict the

potential harmfulness of missense variants. Scores over 0.5 indicate a damaging effect. Mutation Assessor calculates the influence of protein-coding variants based on evolutionary data. Scores below 1.9 suggest a high functional effect, which could significantly disrupt protein function.

Table 2: Information on the 3 variants, including their rsIDs, chromosomal locations, alleles, modified amino acids, and positions of these amino acids.

Sr.No.	Variant ID	Chromosomal location	Alleles	AA	AA coord
1	rs1945280381	7846395	T/C	K/E	322
2	rs746619116	7857506	G/A	R/W	154
3	rs369248365	7865125	G/A/T	R/C	67

Table 3: Score of the VDR gene deleterious variant obtained through multiple common tools.

Sr.No	Variant ID	SIFT	POLYPHEN	CADD	REVEL	MetaLR	Mutation Assessor
1	rs1945280381	Deleterious 0	Possibly damaging 0.901	Likely deleterious 31	Likely disease-causing 0.183	Damaging 0.879	H 0.385
2	rs746619116	Deleterious 0	Possibly damaging 0.717	Likely deleterious 32	Likely disease-causing 0.863	Damaging 0.899	H 0.886
3	rs369248365	Deleterious 0.04	Possibly damaging 0.803	Likely deleterious 31	Likely disease-causing 0.675	Damaging 0.892	H 0.188

4.2 Protein Domains

The vitamin D receptor (VDR) protein, take part in numerous physiological processes, and exhibits a nuclear hormone receptor domain, signifying its role within the nuclear receptor superfamily and its function in DNA binding for transcription regulation, especially upon vitamin D_3 binding (InterPro, 2024). Zinc finger motifs within the VDR suggest robust DNA-binding capabilities, essential for transcriptional regulation of vitamin D_3-responsive genes. Furthermore, a ligand-binding domain is indicative of the VDR protein's ability to bind vitamin D_3, initiating receptor activation and downstream gene expression alterations **Table 4** shows the different domains and the amino acids number of amino acid in each domain.

Table 4: Domains of the VDR protein and the number of amino acids involved in each domain.

Domin Name	**Identifier**	**Amino Acid Range**
Nuclear Hormone Receptor NR1	IPR005234	4-428
VitD_rcpt (Vitamin D Receptor	IPR000324	14-301
Nuclear_hormone_rcpt_lig-bd (Nuclear Hormone Receptor Ligand Binding Domain)	IPR000536	127-423
ZnF_hrmn_rcpt (Zinc Finger Hormone Receptor	IPR001628	21-96
DBD_VDR (DNA Binding Domain of Vitamin D Receptor	IPR024153	16-122
NHR-like_dom_sf (Nuclear Receptor Superfamily	IPR035500	124-423
ZnF_NHR/GATA (Zinc Finger Nuclear Hormone Receptor/GATA	IPR013088	15-124

4.3 VDR Exons

Then to analyze which exonic regions are involved in the coding of amino acid residues of VDR protein, exon coordinates were obtained from Ensembl (ENST00000549336.6), demonstrating that Chromosome 12 on VDR-206 included 10 exons, ranging from 47,841,537 to 47,904,994 on the reverse strand. After opening the VDR gene on NCBI, we were able to access the sequence text viewer by scrolling through the tools section under the Genomic areas heading. Next, we opened the VDR gene sequence and compared it to the ensemble coordinate displayed in the sequence view

at NCBI. This allowed us to find and note our amino acids, as shown in **Figure 11** Exon 3 encodes amino acids to 49, and exon 10 encodes the greatest number of amino acids 85. This analysis was done to determine the number of variations per exon of the VDR gene. The rs746619116 lies inside the Exon 5 which encodes amino acid sequence 93-154. **Table 5** shows the order of the exons in which they code for their amino acid sequences.

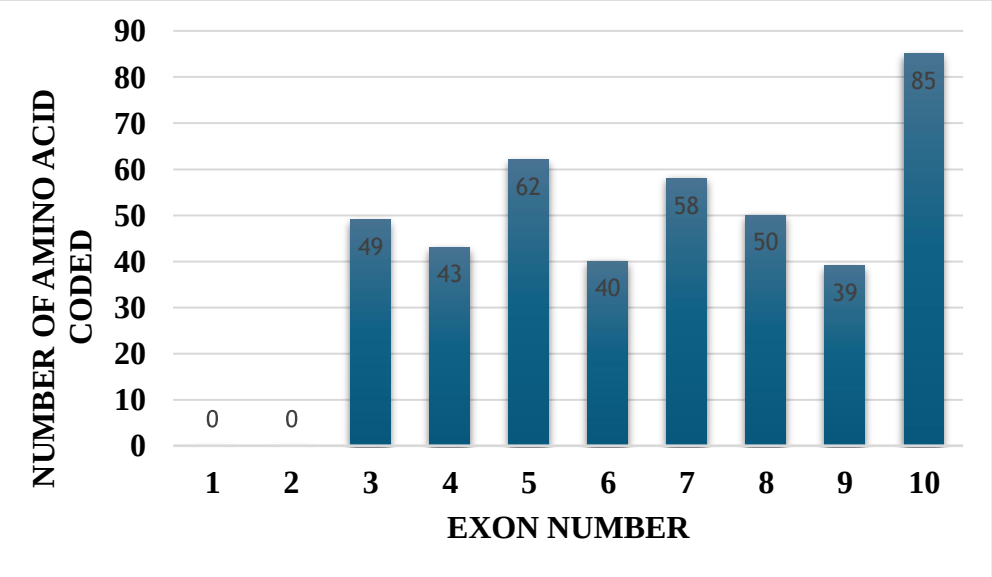

Figure 11: Number of amino acids coded by each exon.

Table 5: Order of exon in which they code for their amino acid sequences.

Exon	AA
1	
2	
3	1-49
4	50-92
5	93-154
6	155-194

7	195-252
8	253-302
9	303-341
10	342-426

4.4 Structure Retrieval and Mutagenesis

The mutant amino acid sequence was introduced into the VDR gene's wild-type structure using the computational tool pyMOL. The VDR gene's 3D structure was first retrieved from alpha folds in the PDB file. It was then opened on the pyMOL program, where the 3D structures of the wild-type and mutant VDR genes were downloaded as indicated. Next, in the external graphical user interface, the Wizard option was selected. The option "Mutagenesis of protein" is included in this selection. This option was displayed in the object control panel, giving the user the option of changing the mutation of an amino acid to match our missense variant (rs746619116). Images generated by PyMOL, we see a detailed 3D rendering of a protein with a highlighted amino acid – Arginine at position 154, which is the point of mutation. This visual cue on the wild-type structure allows us to identify the precise location of the mutation. The corresponding structure, which is the mutant form, shows the result of substituting Arginine with Tryptophan at the same position. The alteration is visible and emphasized, allowing for an assessment of how this single change might alter the protein's conformation and possibly its interaction with other molecules. **Figure 12** and **Figure 13** show the 3D results obtained from pyMOL.

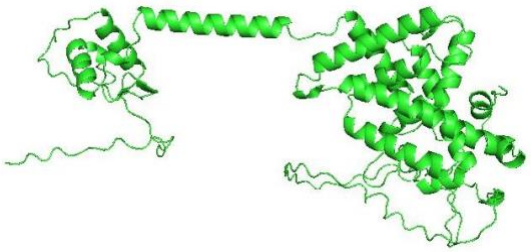

Figure 12: Wild-type image obtained from pyMOL

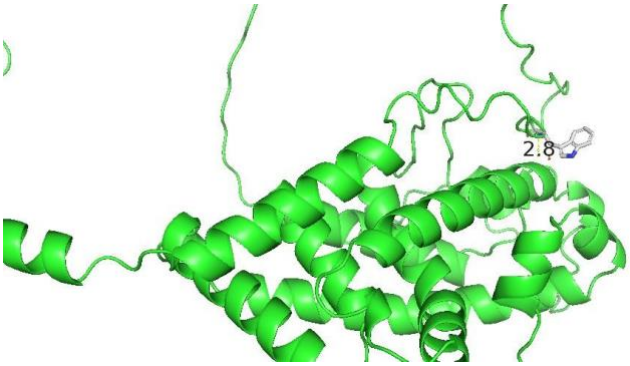

Figure 13: Mutant-type image obtained from pyMOL

4.5 Impact of SNP on Protein Stability

A confidence score of -0.4 from Mupro and a similar tool I-mutant indicated a reduction in protein stability because of the SNP change, suggesting that the SNP was likely to destabilize the protein structure, making it less stable than its natural, wild-type version. Additionally, a ΔΔG value of -0.71 further emphasized this destabilization, indicating a significant decrease in thermodynamic stability, which points to the protein becoming less stable energetically. This negative ΔΔG value meant that the mutated protein was energetically less favored, leading to potential structural

alterations that could impair the protein's functionality. Another tool was used to verify these results: Maestro software provided an analysis of a particular change in the VDR protein, where the amino acid Arginine was switched to Tryptophan at position 154. The predicted change in stability was 0.938, suggesting that this new version of the protein might be more stable than the original. The confidence level of this prediction is quite high, at 0.846, indicating that we can trust this forecast to a good extent. The additional information given by the software highlights that this specific mutation has been found in people and is listed in genetic databases, but its importance for health isn't clear yet. It's been seen in individuals with a certain type of bone disorder, but there isn't enough evidence to confirm that this mutation is the cause. Therefore, while the mutation seems to make the protein more stable, its actual impact on health remains uncertain, and its uncertain.

4.6 Impact of SNP on Protein Structure and Function

The results from Project HOPE indicated a significant mutation where Tryptophan replaces Arginine shown in **Figure 14** at position 154, leading to several critical changes in the amino acids characteristics, including size, charge, and hydrophobicity change from the smaller, positively charged wild-type Arginine to the larger, neutral Tryptophan. This modification from a positive to a neutral charge could interrupt the electrostatic interactions that are essential for the protein's interaction with other molecules, possibly affecting its functions. Moreover, the greater size of Tryptophan may cause structural interruption, creating positional restrictions that could impact the protein's stability and function. Tryptophan's greater hydrophobicity in contrast to Arginine may lead to a decrease in the number of hydrogen bonds, which are vital for maintaining protein structure and function. This increase in hydrophobicity may also negatively impact the correct folding and stability of the protein. The mutation is found within the Vitamin D Receptor's ligand-binding domain, which is part of the Nuclear Hormone Receptor-Like Domain Superfamily. Any interruption in this domain could substantially affect the receptor's ability to bind molecules, possibly modifying essential functions such as DNA binding and molecular transducer activity, with wide-reaching biological effect. With a MetaRNN score of 0.86814785, there is a high likelihood that this mutation is pathogenic. The mutation's presence in a conserved area highlights its potential importance, despite its rarity.

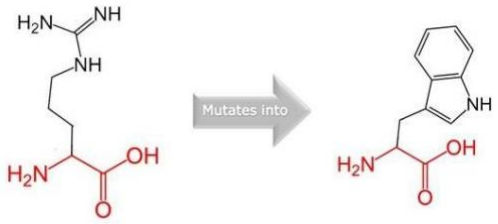

Figure 14: Amino Acid transition.

Another tool DynaMut 2 was used to check the effect of the SNP on the structure of the protein. The protein's wild-type structure is displayed in **Figure 15,** which displays the results from DynaMut2 focusing on the interaction between specific amino acids. Lysine 413 is linked to Aspartate 232 through an ionic bond, represented with orange dashed lines, reflecting the electrostatic interactions between these charged residues. Furthermore, Aspartate 232 forms hydrogen bonds with Proline 228 and Arginine 154, shown with yellow dashed lines, essential for upholding the protein's structural integrity. There is also a hydrogen bond between Arginine 154 and Histidine 229. The Mutant structure is shown in Error! Reference source not found. **.** In the m utant structure, Aspartate 232 (ASP232) still interacts with Proline 228 (PRO228) and Histidine 229 (HIS229), similarly forming hydrogen bonds, as shown by the yellow dashed lines. Similar to the wild type, these interactions aid in stabilizing the protein structure. One significant difference in the mutant structure is the substitution of Tryptophan 154 (TRP154) for Arginine 154 (ARG154). Tryptophan's larger size and different chemical properties from arginine could cause a change in the chemical interactions ultimately altering how the protein interacts with its cellular environment, this substitution could potentially affect the protein's overall stability and function as well as how it interacts with other molecules.

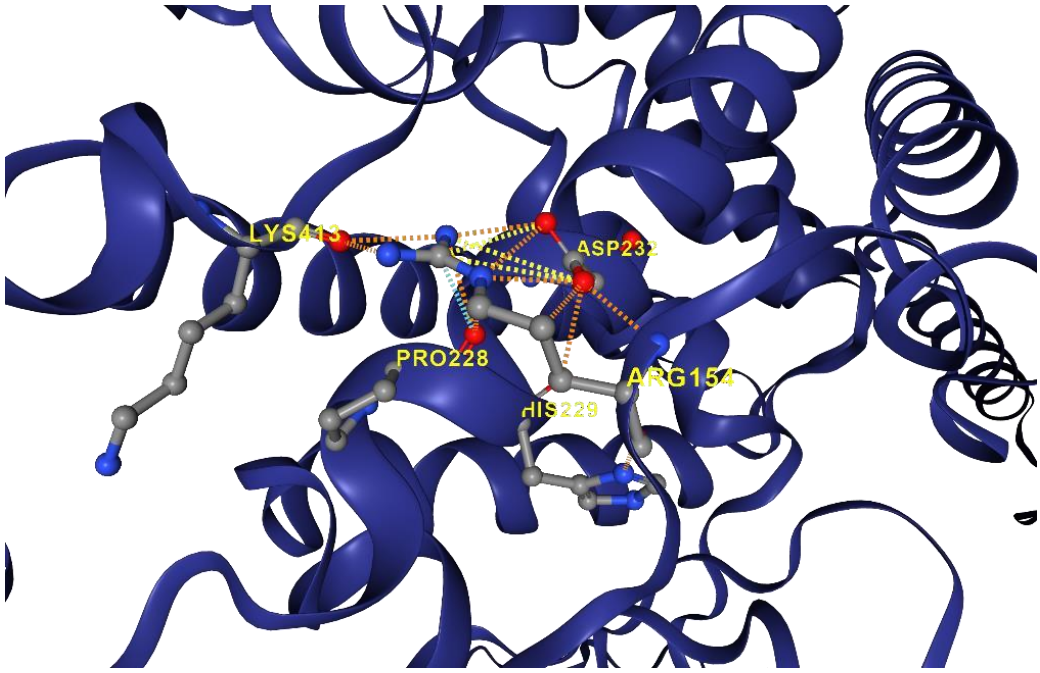

Figure 15: Wild type structure obtained from DynaMut2.

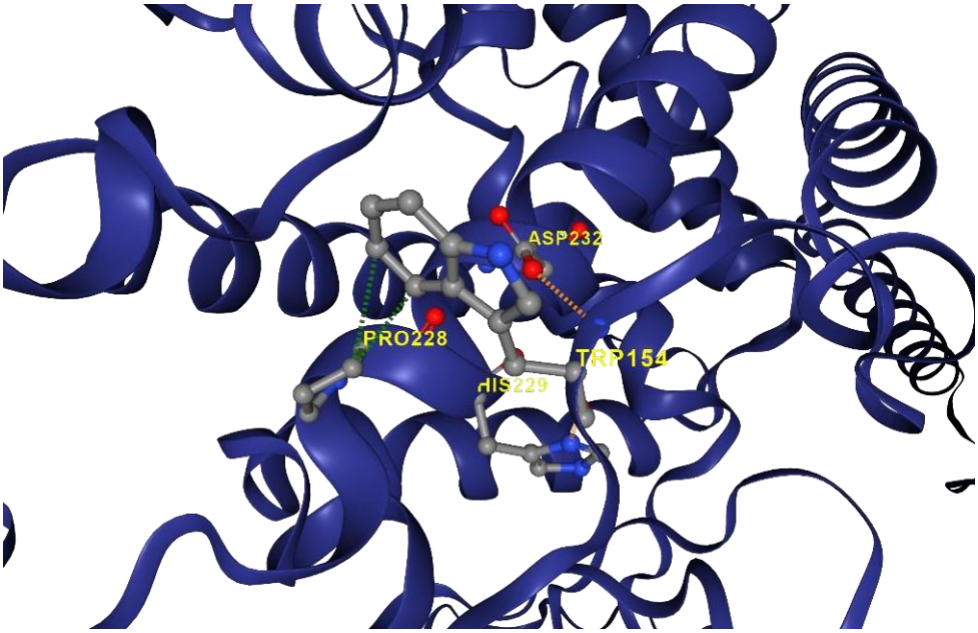

Figure 16: Mutant structure obtained from DynaMut2.

4.7 VDR gene Variant rs746619116 association with Disease

The FATHMM results suggest that this is predicted to be damaging, with a score of -3.93. This negative score often indicates a mutation with significant implications for the protein's function, potentially impacting the health of an individual. The array of conditions listed associates this genetic alteration with various diseases, including several types of cancers and metabolic disorders. Notably, this mutation may also be relevant to HIV, as the VDR gene has been implicated in immune system regulation, which is a key factor in HIV infection and progression. The implication of such a mutation in HIV infection warrants further investigation to understand its role in the disease's pathophysiology and its potential as a biomarker for susceptibility or disease progression. The MutPred2 result for the mutation generates a MutPred2 score of 0.676 suggesting a reasonable likelihood that it could be linked with causing disease, as higher scores translate to greater risks. By decreasing the protein flexibility through the loss of intrinsic disorder with probability of 39% This mutation may cause disease as the flexibility is essential for the normal functioning of the protein Additionally, there is 20% probability that this mutation could cause the deletion of a modification named as ADP-ribosylation at position 158 which is essential for cellular signaling and DNA repair. Furthermore, the results of this tool show that the mutation impacts 2 particular motifs in the protein possibly disrupting how the protein interacts with other molecules in its surroundings and how it carries out its function and possibly leads towards a disease.

4.8 Cellular Localization

The deeploc results showed that the protein is mainly located in the cytoplasm and the nucleus **Table 6** shows the various predicted locations of the protein and the associated probabilities. The graph, in **Figure 17** illustrates conserved amino acid regions and helps identify key motifs that are crucial to the protein's function.

Table 6: Location and probability score of the VDR gene.

Location	Probability
Cytoplasm	0.7995
Nucleus	0.7996

Extracellular	0.0231
Cell Membrane	0.0971
Mitochondria	0.1548
Plastid	0.0049
Endoplasmic Reticulum	0.0800
Lysosome/Vacuole	0.1419
Golgi Apparatus	0.1284
Peroxisome	0.0055

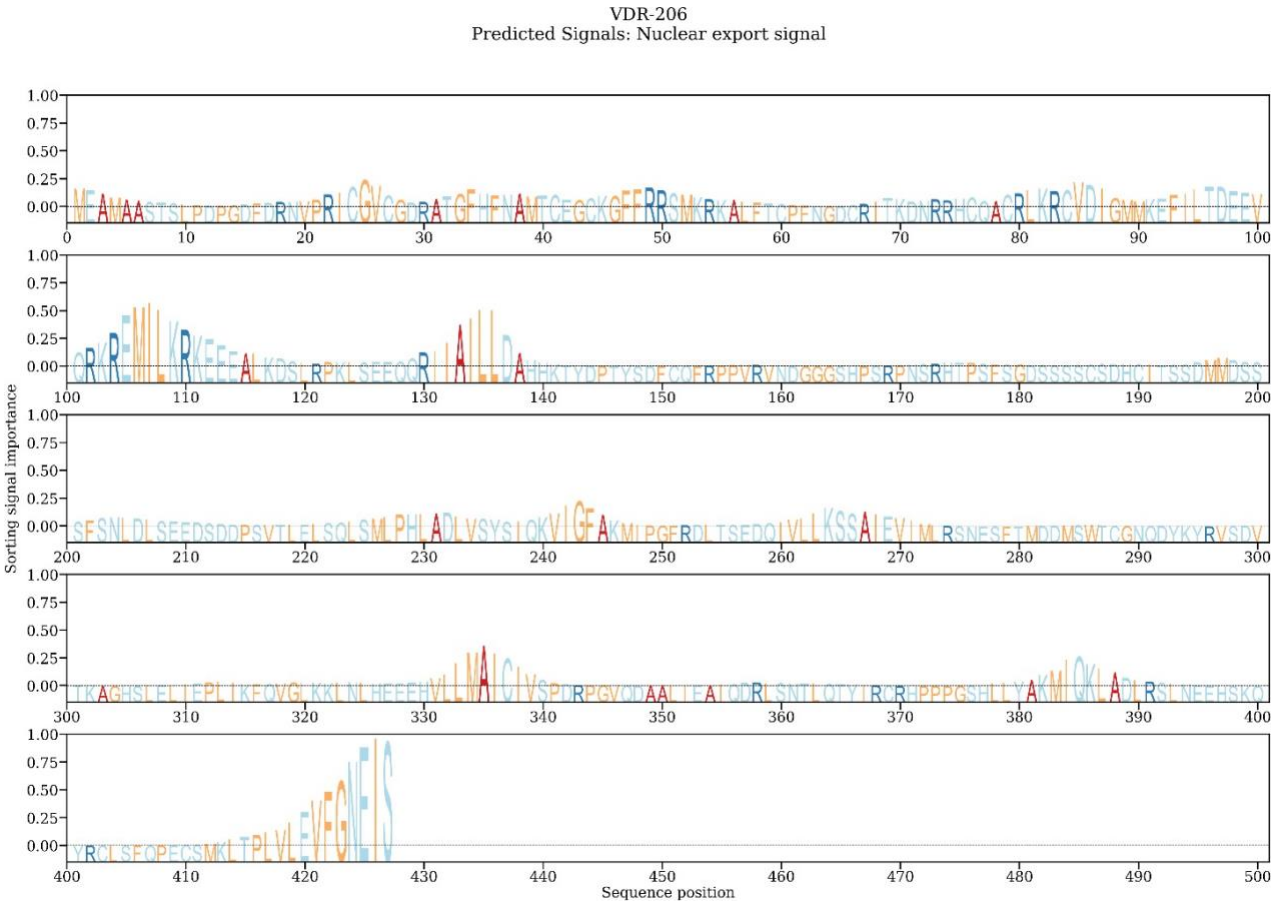

Figure 17: Graph showing conserved amino acids.

The x-axis represents the positions in the amino acid sequence, while the y-axis shows how conserved the positions are across species based on a sorting signal importance value. Taller characters suggest a higher level of conservation. The important regions shown in the graph are amino acid positions 0-100 in this section, highly conserved motifs are shown, is likely important for DNA binding or interactions with other proteins. Additionally amino acid positions 100-200 show further conserved motifs in this region which suggests nuclear receptor functions. Positions 200-30 this area could contain structural elements or ligand-binding sites. Lastly positions 300-400 patterns shown might relate to transcription regulation or other nuclear receptor functions. The analysis, which combines predicted localization and structural motifs, points to a significant regulatory role for this protein.

4.9 Genetic Association of rsID746619116 in HIV Diagnosis

DNA extraction from the samples was performed to analyze the association of the rs74661911 missense variant in these samples, the ARMS-PCR technique was utilized for genotyping purposes. The results, which include the frequencies of the genotypes, odds ratios, relative risks, and p-values for both control and patient samples, are presented in **Table 7.**

Table 7: Frequencies of the genotypes, odds ratios, relative risks, and p-values for both control and patient samples are presented.

	Frequency Distribution		Odd Ratio		Relative Risk		
Genotype	Control	Patient	value	95%CI	Value	95%CI	P Value
CC	44%	30%	1.833	0.773-4.208	1.338	0.898-1.956	0.213
TT	22%	12%	2.068	0.753-6.262	1.377	0.840-1.987	0.286
CT	34%	58%	2.6	0.166-0.852	1.6	0.384-0.913	0.026

Understanding odds ratios (OR) and relative risks (RR) along with their 95% confidence intervals (CIs) and p-values helps identify whether a factor increases or decreases the risk of an outcome in studies. If the OR or RR is above 1, the factor is likely increasing risk, making it a risk factor. If it's below 1, the factor is likely decreasing risk, making it protective. A value of 1 indicates no effect. The significance of these results is confirmed when the 95% CI does not include 1 and the p-value is below 0.05, suggesting that the results are statistically significant. If the CI includes 1, it means the results might be due to chance, indicating no clear association.

Frequency Distribution in the control group 44% have the CC genotype, compared to 30% in the patient group. Odds Ratio (OR) of 1.833 with a 95% Confidence Interval (CI) of 0.773-4.208 This suggests that having the CC genotype is associated with lower odds of being in the patient group, but the association is not statistically significant since the CI includes 1. Relative Risk (RR) of 0.740 with a 95% CI of 0.898-1.956. The relative risk is less than 1 which indicates a potentially protective effect, but the confidence interval is wide and crosses 1, suggesting no clear evidence of a significant effect. P-value is 0.213 this value is greater than 0.05, indicating that the observed difference in genotype frequency between the control and patient groups is not statistically significant. Frequency Distribution of the TT genotype is 22% of the control group and 12% of the patient group have the odds Ratio is 2.068 with a 95% CI of 0.753-6.262. The odds ratio again suggests a lower odd of the disease for the TT genotype, but with a confidence interval that widely crosses 1, it is not statistically significant relative risk of 1.377 with a 95% CI of 0.840 -1.987. This risk ratio indicates a potential risk effect but is not statistically significant due to the confidence interval including 1. P-value is 0.286, which indicates no significant association between the TT genotype and HIV, as it is above the typical P-value level of 0.05.

Frequency Distribution of the CT genotype is more prevalent in the patient group 58% compared to the control group (34%) an odds Ratio of 2.6 with a 95% CI of 0.166-0.852. This indicates a higher odd of being in the patient group associated with the CT genotype. The confidence interval is less than 1, suggesting a potentially significant association. Relative Risk of 1.6 with a 95% CI of 0.384-0. 913. The RR suggests a risk effect of the disease with the CT genotype, the confidence interval does not include 1, indicating that this association is precise with the p-value of 0.026 which is below the threshold suggesting the association is statistically significant. The CC and TT genotypes show a non-significant association with HIV infection due to confidence intervals

crossing 1. The CT genotype, however, demonstrates a statistically significant correlation with HIV risk, indicated by RR and OR above 1 and a significant p-value.

4.10 Genotype Association with Age

Genotype association with age the age of the patients in the Control and Disease groups were both recorded to check the association of age with the genotype present **Table 8** shows the results.

Table 8: Genotype association with age in HIV individuals.

	Frequency Distribution		Odd Ratio		Relative Risk		
Genotype	Below 40	Above 40	value	95%CI	Value	95%CI	P Value
CC	32.4%	33.2%	0.960	0.193-5.601	0.994	0.685-1.231	0.999
TT	8.10%	50%	0.117	0.022-0.684	0.558	0.208-0.932	0.041
CT	59.4%	16.6%	7.333	0.927-89.63	1.275	0.998-1.812	0.081

The results show the TT genotype may serve as a protective factor in older individuals, while the CT genotype could possibly be a risk factor for younger individuals, however more studies are necessary to solidly establish these associations. The CC genotype does not show any significant associations with HIV risk based on age.

The CC genotype frequency Distribution is 32.4% in individuals under 40 years and 33.2% in above 40 years the odds Ratio is 0.960, indicating a slightly lower odds of individuals above 40 having this genotype, however the difference is minimal. With a 95% Confidence Interval (CI) score of 0.193 to 5.601, which is very wide and indicates a high level of uncertainty in the estimate. The relative Risk (RR) value of 0.994 with a 95% CI of 0.685 to 1.231 the P-Value is 0.999, which is not statistically significant as it is above the threshold, suggesting no significant age-related difference in the frequency of this genotype among the patients. The CC genotype does not appear to be a risk or protective factor for HIV.

The TT Genotype has a Frequency Distribution of 8.1% in individuals under 40 years compared to 50% in those above 40 years <Odds Ratio (OR) value is 0.117, which suggests a significantly lower likelihood of younger individuals having this genotype. The 95% Confidence Interval (CI) is between 0.022 and 0.684, which backs the significant odds ratio. The relative risk value is 0.558 with a 95% CI of 0.208 to 0.932 the P-Value is 0.041, which indicates statistical significance. The TT genotype appears to be less common in younger individuals and is more prevalent in older individuals, highlighting its potential as a protective effect against early HIV infection or progression in younger age groups.

Next is the CT genotype with a frequency Distribution of 59.4% in individuals under 40 years and 16.6% in those above 40 years. The Odds Ratio (OR) value is 7.333, suggesting a significantly greater likelihood of younger individuals having this genotype. With a 95% Confidence Interval (CI between 0.927 - 89.63, which is wide, it indicates a higher prevalence of this genotype in the younger age group. The Relative Risk (RR) value is 1.275 with a 95% CI of 0.998 to 1. 812.The P-Value is 0.081, which is not statistically significant but close to the threshold, hinting toward a trend. While not statistically significant, the higher prevalence of the CT genotype in younger individuals may suggest it is a risk factor for HIV for this genotype. However, further research with more individuals involved is needed to confirm this.

4.11 Vitamin D levels and Platelet Count Association with Genotype in HIV patients

Vitamin D levels and platelet counts were measured in HIV-positive patients, as shown below. Error! Reference source not found. displays the average vitamin D levels for each genotype, Having t he CT genotype is associated with having greater vitamin D levels as compared to other genotypes. The difference between the TT and the CC genotype is not significant. This data suggests that CT genotype can greatly increase vitamin D levels in HIV-positive patients. Error! Reference source n ot found. illustrates the average platelet counts for each genotype among these patients The CC genotype has the lowest average platelet count of around 40,000 per µl while the TT genotype shows a moderately higher average platelet count of 80 000 per µl of however the CT genotype displays a greatest average platelet count of around 160000 per µl which is significantly higher than the other genotypes this indicates a potential association of having CT genotype and increased platelet count in HIV patients. Additionally, some individuals in our HIV-positive group also had thalassemia shows the average thalassemia cases for each genotype the TT genotype is associated with the lowest number of thalassemia cases in HIV-positive patients while the CT genotype is associated with significantly higher thalassemia cases which indicates that having the CT genotype in HIV patients increases the risk of developing thalassemia.

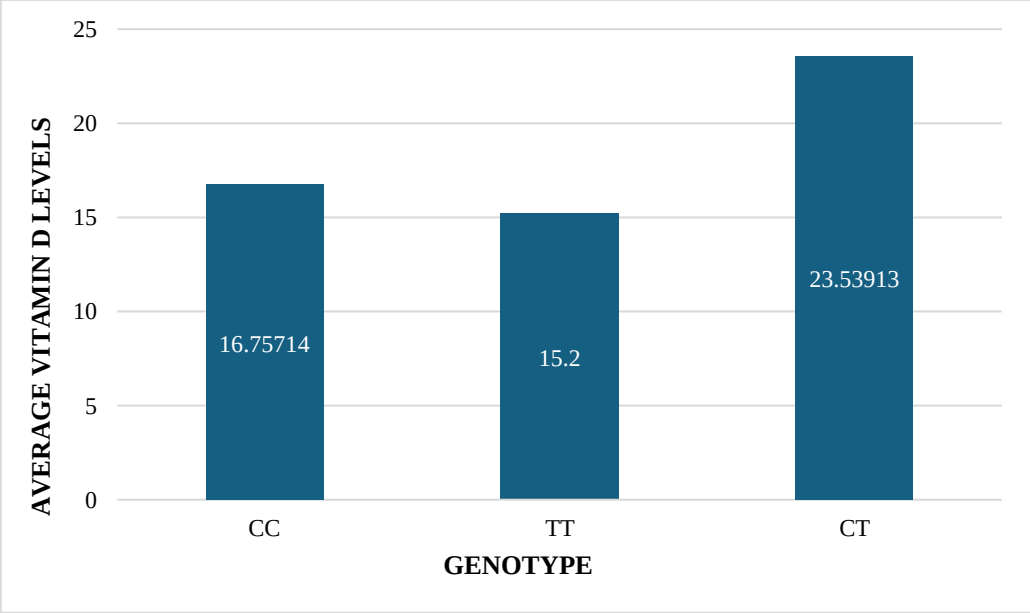

Figure 18: Average vitamin D levels for each genotype.

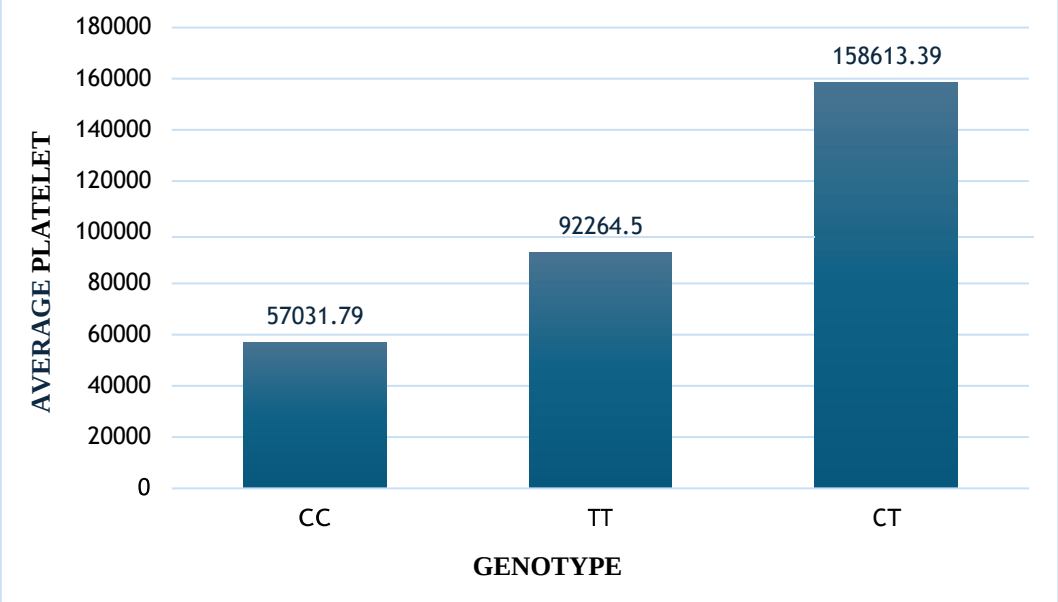

Figure 19: Average platelet counts for each genotype.

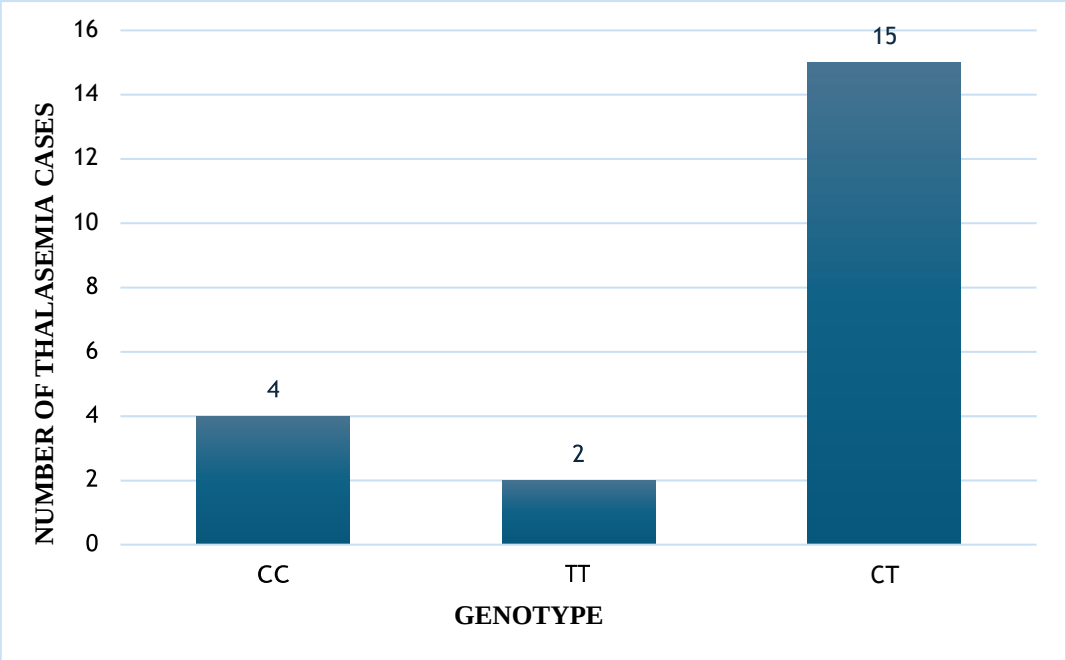

Figure 20: Average thalassemia cases for each genotype.

Chapter 5: Discussion

HIV significantly compromises the immune system by specifically targeting CD4 cells, which leads to increased susceptibility to various opportunistic infections and can eventually progress to AIDS if left untreated. Global statistics indicate that approximately 85.6 million people have been infected with HIV to date, resulting in around 40.4 million fatalities due to HIV-related complications. Currently, the worldwide HIV-positive population stands at about 39.0 million. The burden of HIV is notably severe in the African Region, where approximately 3.2% of the adult population is affected. Efforts to curb this epidemic included widespread testing, with about 190 million people receiving HIV testing results in 2018 alone. Investigating the correlation between vitamin D deficiency or genetic anomalies in the Vitamin D Receptor (VDR) gene and the progression of HIV could pave the way for innovative diagnostic techniques. Vitamin D and its receptor, VDR, along with the active hormone form, 1α,25-dihydroxyvitamin D_3, are integral to immune function. Exploring how these elements interact could unravel new aspects of HIV pathogenesis and potentially lead to the development of novel diagnostic tools to track disease progression and identify individuals at elevated risk of developing severe disease states. Research in this area includes numerous studies linking specific VDR gene variants to HIV outcomes. For example, variations such as the Bsm-I A allele have been associated with faster disease progression in HIV-infected children older than two years. A separate study focused on a black Southern African cohort examined the prevalence of VDR polymorphisms (Apa-I, Bsm-I, FokI, and TaqI) among HIV-positive individuals compared to HIV-negative peers. Here, notable findings included the Apa-1 genotype showing varied distribution patterns based on HIV status in both dominant and co-dominant genetic models. Moreover, another significant study highlighted the protective nature of the T allele at the VDR rs2228570 SNP against AIDS progression in individuals who have not undergone antiretroviral therapy. The objective of this study was to investigate the influence of the missense variant rs746619116 in the VDR gene on the structure and function of the protein, specifically focusing on its flexibility, utilizing multiple tools to determine if this variant is linked with HIV, along with its pathological and clinical impacts.

Additionally, the association between the VDR protein and HIV was explored through genotype analysis in both patient and control groups. A dataset of 28,458 missense variants from the VDR gene was obtained from the Ensemble database. To assess the pathogenicity of these variants, six

computational tools were employed: Polyphen2, SIFT, REVEL, MetaLR, Mutation Assessor, and CADD. Three pathogenic missense mutations were found by the analysis: K322E, R154W, and R67C. The mutation with the C/T allelic variation, rsID rs746619116, that converts arginine to tryptophan at position 154 in the VDR protein sequence was singled out for special attention. A thorough in-silico analysis, as well as genotype research, were carried out with this mutation in the VDR protein. The Project HOPE analysis revealed that the substitution of tryptophan for arginine at position 154 has a substantial impact on the DNA-binding domain of the VDR protein. Important properties like size, charge, and hydrophobicity are impacted by this alteration, which may compromise the protein's ability to regulate gene expression and have an effect on vital physiological functions like immune modulation and calcium regulation. PyMOL visual analyses demonstrate that this The VDR protein's structural integrity and interaction potential, leading to notable conformational changes that may compromise its ability to perform normal functions. This mutation has a destabilization score of -0.71 according to DynaMut 2, which suggests a reduction in the thermodynamic stability of the protein that may impair its functionality and increase the risk of disease. Furthermore, FATHMM gives the mutation a score of -3.93, indicating that it may have a significant negative impact on protein function and may be linked to diseases like cancer, metabolic disorders, and altered immune systems that affect HIV infection. Conversely, Maestro suggests a potential increase in protein stability with a score of 0.938. This introduces a nuanced view of the mutation's impact, suggesting that under certain conditions, the mutation may offer some structural advantages, although the overall implications for health are still uncertain. MutPred2 evaluation, with a score of 0.676, supports the notion that this mutation is likely to be pathogenic.

In analyzing the association between the CC, TT, and CT genotypes with HIV infection, it's clear that the results vary among different genotypes. The CC genotype is more frequent in the control group compared to the patient group, suggesting a potential protective effect against HIV. However, the odds ratio of 1.833 and the relative risk of 0.740 aren't statistically significant, as indicated by confidence intervals that cross 1 and a p-value above 0.05. This means that the protective effect can't be firmly established. Similarly, the TT genotype is also more prevalent in the control group than in the patient group, suggesting a reduced potential of infection. , the odds ratio of 2.068 and the relative risk of 1.377 lack statistical significance due to the wide confidence intervals crossing 1. The p-value of 0.286 further highlights the absence of a clear association. However, the CT

genotype is significantly more common among HIV-positive individuals compared to the control group. The odds ratio value of 2.6 and relative risk of1.6 indicate a risk associated with this genotype, with confidence intervals that don't cross 1 and a p-value of 0.026. This result implies a statistically significant correlation between the CT genotype and a higher susceptibility to HIV infection. Overall, while the CC and TT genotypes don't show clear associations with HIV infection, the CT genotype appears to be significantly linked to an increased risk. The frequency distribution of the CC genotype across age groups, the slight difference between individuals younger and older than 40 is minimal, with both percentages ranging around 32-33%. The odds ratio and relative risk don't suggest a significant variation between these groups, and the wide confidence intervals indicate uncertainty, meaning this genotype isn't linked to age-related HIV risk. The TT genotype, however, presents a more notable difference. It's less common in younger individuals than in those above 40. The low odds ratio and statistically significant confidence intervals support this claim, suggesting the TT genotype may have a protective role in delaying or preventing HIV infection or progression among younger individuals. In contrast to its frequency in older age groups, the CT genotype is more common in people under 40. A statistical analysis indicates that there is a greater probability of this genotype being present in younger individuals, which may indicate that it is a risk factor for HIV in this population. Although the p-value is not significant and the confidence interval is wide, there is an apparent pattern that points to a potential correlation.

Further extensive research is necessary to validate these initial results. In summary, younger people may be at risk for HIV if they have the CT genotype, whereas older individuals may benefit from the protective effects of the TT genotype. However, there is no discernible relationship between the age-related HIV risk and the CC genotype. Further investigation into these connections through in-depth studies may have an impact on HIV. Additionally, results revealed that HIV individuals with the CT genotype typically exhibit higher vitamin D levels compared to those with the CC and TT genotypes, suggesting the CT genotype may increase vitamin D absorption or metabolism in HIV-positive individuals. Platelet counts in HIV patients also vary significantly among the genotypes, with the CT genotype associated with the highest counts, around 160,000 per µl, much higher than those in individuals with CC and TT genotypes. This suggests a potential genetic influence on blood health, where the CT genotype could be linked to higher platelet production or stability in HIV patients. Furthermore, these genotypes are more complex due to the prevalence of thalassemia. In addition to having higher platelet counts and vitamin D levels, the CT genotype also exhibits a

higher prevalence of thalassemia, indicating a genetic predisposition to this illness in addition to HIV.

Chapter 6: Conclusion and Future Prospects

The VDR gene variant rs746619116 has been studied, and although the results are preliminary, they show that the gene may have applications in HIV diagnosis. The structure and function of the VDR protein may change as a result of the mutation, according to in-silico analysis, impacting critical immune system interactions that are essential for HIV management. Experimental studies display a likely risk effect of CT genotype. However, these results require additional validation therefore, additional research engaging diverse populations is important to clarify the variant's role and validate its effectiveness as an HIV diagnostic marker. In addition to extending the study population, in-depth encounters with HIV patients in clinical settings might yield useful insights. Gathering data on their HIV diagnosis date, current drug regimen (with an emphasis on essential antiretroviral medicines), CD4 count, and any treatment-related side effects will help us better understand HIV progression and outcomes. With an increased sample size, we can further look into a critical features research can be done in comparing people on ART (Antiretroviral Therapy) to those who are not on treatment. A thorough study on the influence of various drug regimens on HIV progression is required. It will help to better understand the complex interplay between genetic predispositions, treatment efficacy, and disease progression by examining patterns and relationships among VDR variations, HIV status, ART drug, and other parameters.

Chapter 7: References

A D Frankel, J. A. (1998). HIV-1: fifteen proteins and an RNA. *Annual review of biochemistry*, 1-25. doi: 10.1146/annurev.biochem.67.1.1

Amaran Moodley, M. M. (2013, November). *Vitamin D Related Host Genetic Variants Alter HIV Disease Progression in Children.* Retrieved from NIH Public Access: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3811908/

Andrew Collins, X. K. (2012). Primer1: Primer Design Web Service for Tetra-Primer ARMS-PCR. *The Open Bioinformatics Journal, 6*(1), 55-58. doi:10.2174/1875036201206010055

Boris Reva, Y. A. (2011). Predicting the functional impact of protein mutations: application to cancer genomics. *Nucleic acids research, 39*(12), e118. doi:10.1093/nar/gkr407

Capriotti E, F. P. (2005). I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. *Nucleic Acids Res.*, doi: 10.1093/nar/gki375. PMID: 15980478; PMCID: PMC1160136.

Carlos H.M. Rodrigues, D. E. (2020). DynaMut2: Assessing changes in stability and flexibility upon single and multiple point missense mutations. *Protein Science*. Retrieved from https://doi.org/10.1002/pro.3942

Chang Li, D. Z. (2022). MetaRNN: differentiating rare pathogenic and rare benign missense SNVs and InDels using deep learning. *Genome medicine, 14*(1), 115. doi:10.1186/s13073-022-01120-z

Chang, W. C. (n.d.). Retrieved from Primer 1.

Dulbecco, R. (1988). Virology (2nd edition). In H. S. Renato Dulbecco, *Virology (2nd edition)* (pp. 555-381). Philadelphia: J. B. Lippincott company.

Emidio Capriotti, P. F. (2005). I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. *Nucleic acids research*, W306-310. doi:10.1093/nar/gki375

Fathi, N. &. (2019). *Role of vitamin D and vitamin D receptor (VDR) in oral cancer.* Tabriz, Iran: Biomedicine & Pharmacotherapy.

Fergal J Martin, M. R.-O. (2023). Retrieved from Ensembl: https://doi.org/10.1093/nar/gkac958

Gema Nieto, Y. B. (2004). Association between AIDS disease progression rates and the Fok-I polymorphism of the VDR gene in a cohort of HIV-1 seropositive patients. *The Journal of Steroid Biochemistry and Molecular Biology*, 199-207. Retrieved from https://doi.org/10.1016/j.jsbmb.2004.03.086.

Geng, C. (2021). *What are the different types of HIV tests and how quick and accurate are they?* Retrieved from MedicalNewsToday: https://www.medicalnewstoday.com/articles/types-of-hiv-tests

Hashem A Shihab, J. G. (2013). Predicting the Functional, Molecular, and Phenotypic Consequences of Amino Acid Substitutions using Hidden Markov Models. *Human mutation, 34*(1), 57-65. doi:10.1002/humu.22225

Hendro Hendro, E. S. (2020). Distribution of VDR Gene Polymorphisms Bsm-I rs1544410 and Apa-I rs7975232 among HIV/AIDS Patients from West Java. *Majalah Kedokteran Bandung*, 40-44. doi:10.15395/mkb.v52n1.1680.

HIV Testing. (2022). Retrieved from Centers for Disease Control and Prevention: https://www.cdc.gov/hiv/testing/index.

Ivan Adzhubei, D. M. (2013). Predicting functional effect of human missense mutations using PolyPhen-2. *Current protocols in human genetics*. doi:10.1002/0471142905.hg0720s76

K Mullis, F. F. (1986). Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harbor symposia on quantitative biology*. doi: 10.1101/sqb.1986.051.01.032

Laimer J, H. H. (2015 Apr 16). MAESTRO--multi agent stability prediction upon point mutations. *BMC Bioinformatics.* doi:10.1186/s12859-015-0548-6.

MacDonald, D. R. (2004). Vitamin D Receptor. In W. J. Lane, *Encyclopedia of Biological Chemistry* (pp. 378-382). Case Western Reserve University, Cleveland, Ohio, USA.

Martin Kircher, D. M. (2014). A general framework for estimating the relative pathogenicity of human genetic variants. *Nature genetics, 46*(3), 310-315. doi:10.1038/ng.2892

MedicalNewsToday. (n.d.). Retrieved from Differences between HIV-1 and HIV-2: https://www.medicalnewstoday.com/articles/323893

Meropi Dimitriadou, A. C.-M. (2011). Fok-I gene polymorphism of vitamin D receptor in patients with beta-thalassemia major and its effect on vitamin D status. *Hematology, 16*(1), 54-58. doi:10.1179/102453311X12902908411878

Muhammad Aizaz, F. A. (2023, July 28). *Health Sci Rep*. Retrieved from Alarming rise in HIV cases in Pakistan: Challenges and future recommendations at hand: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10375546/#:~:text=According%20to%20the%20Figure%202,below%2015%20years%20are%204600.

Nicola Gianotti 1, G. M.-L. ((1999),). Brief Report: Drop in CD4+ Counts Below 200 Cells/μL After Reaching (or Starting From) Values Higher than 350 Cells/μL in HIV-Infected Patients With Virological Suppression. *Journal of acquired immune deficiency syndromes*, 417–422.

Nilah M. Ioannidis, J. H.-A. (2016). REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. *American journal of human genetics, 99*(4), 877-885. doi:10.1016/j.ajhg.2016.08.016

Norman, A. W. (2006). Minireview: Vitamin D Receptor: New Assignments for an already busy receptor. *Endocrinology, 147*(12), 5542–5548. doi:10.1210/en.2006-0946

Olga Shcherbatova, D. G. (2020). Modeling of the HIV-1 Life Cycle in Productively Infected Cells to Predict Novel Therapeutic Targets. *Basel, Switzerland, 9*(4), 255. doi: 10.3390/pathogens9040255

Organization, W. H. (2022). *HIV*. Retrieved from World Health Organization: https://www.who.int/data/gho/data/themes/hiv-aids/hiv-aids

Organization, W. H. (2024, May 2). *HIV/AIDS*. Retrieved from World Health Organization: https://www.who.int/health-topics/hiv-aids

Pakistan's Sindh grapples with outbreak of HIV infections. (2019, 7 11). Retrieved from Arab News Pakistan: https://www.arabnews.pk/node/1523866/world

Pauline C Ng, S. H. (2003). SIFT: Predicting amino acid changes that affect protein function. *Nucleic acids research, 31*(13), 3812-3814. doi:10.1093/nar/gkg509

Pejaver V, U. J.-M. (2020). Inferring the molecular and phenotypic impact of amino acid variants with MutPred2. *Nat. Commun.* Retrieved from https://doi.org/10.1038/s41467-020-19669-x

R Carroll, L. M. (1991). Identification of lentivirus tat functional domains through generation of equine infectious anemia virus/human immunodeficiency virus type 1 tat gene chimeras. *Journal of virology, 65*(7), 3460–3467. doi: 10.1128/jvi.65.7.3460- 3467.1991 https://doi.org/10.1128/JVI.65.7.3460-3467.1991

R Swanstrom, J. W. (1997). Synthesis, Assembly, and Processing of Viral Proteins. In *Retroviruses. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 1997.*

Rachel S. Kamgaing, Y. B. (2023). Challenges faced by the HIV testing system in low- and middle-income countries. *African journal of laboratory medicine, 12*(11), 1974. doi: 10.4102/ajlm.v12i1.1974

Rogers MF, S. H. (2018). FATHMM-XF: accurate prediction of pathogenic point mutations via extended features. *Bioinformatics.*, 34(3):511-513. doi:10.1093/bioinformatics/btx536.

Sarudzai P. Showa, F. N.-M. (2019). On the efficiency of HIV transmission: Insights through discrete time HIV models. *PLOS ONE, 14*(9), e0222574. Retrieved from https://doi.org/10.1371/journal.pone.0222574

Schrödinger, L. (2024, May 2). *PyMOL (Version 2.5.8)*. Retrieved from https://pymol.org

Shcherbatova, O. G. (2020). Modeling of the HIV-1 Life Cycle in Productively Infected Cells to Predict Novel Therapeutic Targets. Pathogens. *Basel, Switzerland*, 255.

Stephen A. Spector, M. (2011). *Vitamin D and HIV: letting the sun shine in.* San Diego: Top Antivir Med.

Tamara Elliott, E. J. (2019). Challenges of HIV diagnosis and management in the context of pre-exposure prophylaxis (PrEP), post-exposure prophylaxis (PEP), test and start and acute HIV

infection: a scoping review. *Journal of the International AIDS Society, 22*(12), e25419. doi:10.1002/jia2.25419

Thomas F. Monaghan, S. N. (2021). Foundational Statistical Principles in Medical Research: A Tutorial on Odds Ratios, Relative Risk, Absolute Risk, and Number Needed to Treat. *Int. J. Environ. Res. Public Health, 8*(11), 5669. doi: 10.3390/ijerph18115669

van Heuvel, Y. S. (2022). *Infectious RNA: Human Immunodeficiency Virus (HIV) Biology, Therapeutic Intervention, and the Quest for a Vaccine*. Retrieved from https://doi.org/10.3390/toxins14020138

Varadi, M. e. (2023). AlphaFold Protein Structure Database in 2024: Providing structure coverage for over 214 million protein sequences. *Nucleic Acids Research, gkad1011 (2023). DOI: 10.1093/nar/gkad1011* .

Venselaar, H. t. (2010). Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces. *BMC Bioinformatics 11*. doi: 10.1186/1471-2105-11-548

Vikas Pejaver, J. U.-M.-J. (2020). Inferring the molecular and phenotypic impact of amino acid variants with MutPred2. *Nature communications, 11*(1), 5918. doi:10.1038/s41467-020-19669-x

Vineet Thumuluri, J. J. (2022). DeepLoc 2.0: multi-label subcellular localization prediction using protein language models. *Nucleic Acids Research*, W228–W234. Retrieved from https://doi.org/10.1093/nar/gkac278

Watson, S. (2023). *HIV Testing*. Retrieved from WebMD: https://www.webmd.com/hiv-aids/hiv-aids-screening

Wilber, A. &. (2014). *Gene Therapy for Severe Hemoglobin Disorders.* Illinois, USA: Electronic Library of Science (eLS).

Yasemin van Heuvel, S. S. (2022). Infectious RNA: Human Immunodeficiency Virus (HIV) Biology, Therapeutic Intervention, and the Quest for a Vaccine. *Toxins, 14*(2), 138. doi:https://doi.org/10.3390/toxins14020138

Ye S, D. S. (2001). An efficient procedure for genotyping single nucleotide polymorphisms. *Nucleic Acids Res.* doi:10.1093/nar/29.17.e88.