HAMAD BIN KHALIFA UNIVERSITY

COLLEGE OF HEALTH AND LIFE SCIENCES

GENOME-WIDE ASSOCIATION STUDY OF VITAMIN D DEFICIENCY IN THE MIDDLE EAST WITH A RELEVANT CHARACTERIZATION OF THE NOVEL *SDR42E1* GENE

BY

NAGHAM NAFIZ AHMAD HENDI

A Dissertation Submitted to the Faculty of

COLLEGE OF HEALTH AND LIFE SCIENCES

In Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

April 2023

COMMITTEE

The members of the Committee approve the dissertation of Nagham Nafiz Ahmad Hindi defended on date April 6th, 2023

DocuSigned by:
Georges Nemer
EDD5ABACF508433...

Georges Nemer, Ph.D.
Dissertation Supervisor

DocuSigned by:
Omar Albagha
2C03E73191FD452...

Omar Al-bagha, Ph.D.
Committee Member

DocuSigned by:
Johan Ericsson
C378302C629C45E...

Per Johan M. Ericsson, Ph.D.
Committee Member

DocuSigned by:
Ayman Al Haj Zen
853FDA3D453F456...

Ayman Alhaj Zen, Ph.D.
Committee Member

DocuSigned by:
Rami Mahfouz
FE5C856F6422404...

Rami Mahfouz, MD
Committee Member

Approved:

DocuSigned by:
Georges Nemer
EDD5ABACF508433...

Dr. Georges Nemer, Dean, College of Health and Life Sciences

ABSTRACT

Introduction: Epidemiological studies have revealed that Middle Eastern countries have the highest incidence of Vitamin D deficiency with severe complications. However, the impact of Vitamin D polymorphisms and the performance of polygenic models have been studied primarily in European populations, with little knowledge in the Middle Eastern. A nonsense variant in the uncharacterized *SDR42E1* gene has been identified recently as a potential contributor to Vitamin D deficiency through genomic research.

Methods: I conducted the first genome-wide association study to identify genetic determinants of Vitamin D levels in Middle Eastern populations using whole-genome and whole-exome sequencing approaches in 6,047 and 199 discovery subjects from Qatar and Lebanon, respectively. I also functionally and structurally characterized the novel *SDR42E1* by generating stable CRISPR/Cas9-mediated genome editing in the selected HaCat and HCT116 human cell models.

Results: I discovered a novel variant, rs2298850 (*P*-value = 1.71×10^{-08}, effect size (Beta) = -0.1285), in a known locus of the group-specific component gene (*GC*) in the Qatari population. I confirmed the association of Vitamin D to several variants, including rs11723621 (*P*-value = 1.93×10^{-08}, Beta = -0.12574) and rs4588 (*P*-value = 8.06×10^{-08}, Beta = -0.1188) in the *GC*. I further identified a novel suggestive variant, rs141064014 on chromosome 7 in the *MGAM* gene (*P-value* of 4.40×10^{-06}) and rs7036592, on chromosome 9 in the *PHF2* gene (*P-value* of 8.43×10^{-06}). A GWAS meta-analysis combining results from the previous European data and Qatari cohort identified novel variants in known loci, including rs67609747 and rs1945603 on chromosome 11. Many variants were replicated through combining elderly Lebanese data and the largest European GWAS from the UK Biobank, including rs2725405 on chromosome 17 in the *SLC38A10* gene (P-value of 3.73 x 10^{-08}). Finally, a low predictive performance of European ancestry-derived polygenic scores was observed when applied to the Middle East individuals.

I determined a cytoplasmic localization of SDR42E1 protein in the cutaneous HaCat and intestinal HCT116 cells. Significant gene associations between the *SDR42E1* and genes involved in Vitamin D pathways were identified, including alkaline phosphatase, placental type (*ALPP*), ATP-binding cassette C1 (*ABCC1*), solute carrier 7A5 (*SLC7A5*). Gene regulators of cellular senescence and cancer prognosis were found to be significantly affected after the knockout and knockout of *SDR42E1* in HaCat and HCT116 cells. Significant alterations in Vitamin D metabolites, including 24R-24,25-Dihydroxyvitamin D, and lipid membrane components, including phosphatidylcholine, were observed in the absence of *SDR42E1* from the HaCat cells. Cellular viability also decreased significantly after the knockout of *SDR42E1* in the HCT116 cells.

Conclusion: These results emphasize the diversity in the genetic architecture and its impact on preventive and precision medicine across different populations. My findings offer novel perspectives on the physiological mechanisms and genetic factors contributing to the variation of Vitamin D levels in Middle Eastern populations. The comprehensive understanding of the molecular mechanisms underlying Vitamin D metabolism and associated health conditions garnered from my study of the novel *SDR42E1*, and its variant constitutes a foundation for future research and translational applications in clinical precision medicine.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

Abbreviation	Explanation
1,25(OH)2D	1,25-DihydroxyVitamin D
25(OH)D	25-HydroxyVitamin D
ABCA1	Adenosine 5′-Triphosphate-Binding Cassette, Subfamily A, Member 1
ABCB1	Adenosine 5′-Triphosphate-Binding Cassette, Subfamily B, Member 1
ACTE1P	Actin Epsilon 1 Pseudogene
AMDHD1	Amidohydrolase Domain Containing 1
APOB	Apolipoprotein B
APOC1	Apolipoprotein C1
APOE	Apolipoprotein E
ATP	Adenosine 5′-Triphosphate
Cas9	CRISPR-associated Protein 9
CETP	Cholesteryl Ester Transfer Protein
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CVD	Cardiovascular Disease
CYP24A1	Cytochrome P450 Family 24, Subfamily A, Member 1
CYP27A1	Cytochrome P450 Family 27, Subfamily A, Member 1
CYP27B1	Cytochrome P450 Family 27, Subfamily B, Member 1
CYP2J2	Cytochrome P450 Family 2, Subfamily J, Member 2
CYP2R1	Cytochrome P450 Family 2, Subfamily R, Member 2
CYP3A4	Cytochrome P450 Family 3, Subfamily A, Member 4
DHC	Dehydrocholesterol
DHCR7	7-Dehydrocholesterol Reductase
DNA	Deoxyribonucleic Acid

EBP	Emopamil Binding Protein
ExWAS	Exome-based Genome-wide Association Study
FGF23	Fibroblast Growth Factor 23
FLG-AS1	Filaggrin And Keratinocyte-Associated 1
FOXA2	Forkhead Box A2
GC	Group-Specific Component
GDP	Guanosine Diphosphate
GWAS	Genome-Wide Association Studies
HSD17B11	Hydroxysteroid-17-Beta-Dehydrogenase 11
HSD3	Hydroxysteroid 17-Beta Dehydrogenase 3
HSD3B2	3-Beta-Hydroxy-Delta 2-Steroid Dehydrogenase
HSD3B5	3-Beta-Hydroxy-Delta 5-Steroid Dehydrogenase
HSDs	Hydroxysteroid Dehydrogenases
HSPG2	Heparan Sulfate Proteoglycan 2
IU	International Units
KIF4B	Kinesin Family Member 4B
LDLR	Low-Density Lipoprotein Receptor
LIPC	Lipase C
LIPG	Lipase G
mRNA	Messenger Ribonucleic Acid
NAD	Nicotinamide Adenine Dinucleotide
NADSYN1	Nicotinamide Adenine Dinucleotide Synthetase-1
NCK	Nck Adaptor Protein 1
OH	Hydroxy Group
PADI1	Peptidyl Arginine Deiminase 1

PCSK9	Proprotein Convertase Subtilisin/Kexin Type 9
PRS	Polygenic Risk Score
PTH	Parathyroid Hormone
RNA	Ribonucleic Acid
RXR	Retinoid-X Receptor
SDR	Short-Chain Dehydrogenase/Reductase
SDR3E	Short-Chain Dehydrogenase/Reductase Family 3E
SDR42E1	Short-Chain Dehydrogenase/Reductase Family 42E, Member 1
SDR4E	Short-Chain Dehydrogenase/Reductase Family 4E
SEC23A	Sec23 Homolog A
SERPINB11	Serine Proteinase Inhibitor B11
SNP	Single Nucleotide Polymorphisms
SPF	Sun Protection Factor
SSTR4	Somatostatin Receptor 4
SULT2A1	Sulfotransferase Family 2A1
TINK	TRAF2 And NCK Interacting Kinase
TRAF2	TNF Receptor-Associated Factor 2
TXNIP	Thioredoxin-Interacting Protein
UGT1A5	UDP-Glucuronosyltransferase 1 Family, Polypeptide A5
UK	United Kingdom
US	United States
UV	Ultraviolet
VDBP	Vitamin D-Binding Protein
VDR	Vitamin D Receptor
VDREs	Vitamin D Response Elements

DEDICATION

To my beloved family,

Thank you for your unwavering support and for believing in me. Your love and prayers have been a constant inspiration throughout my academic pursuit. I dedicate this achievement to you with immense gratitude.

ACKNOWLEDGMENTS

I praise Allah, the Almighty, for His blessing and guidance throughout my study.

I would like to express my sincere gratitude to my Ph.D. supervisor, Dr. Georges M. Nemer, for his invaluable guidance and insightful discussions throughout my doctoral journey.

I am thankful to my dissertation committee members for their helpful feedback, and to our collaborators, Dr. Omar Albagha, Dr. Per Johan M. Ericsson, and Dr. Ghada El-Hajj Fuleihan, for their contributions. I also thank the continued assistance of CHLS members.

Finally, I would like to sincerely thank my family for their instrumental encouragement.

CHAPTER 1: INTRODUCTION

1.1 Vitamin D

Vitamin D is a vital fat-soluble nutrient that plays a crucial role in maintaining bone mineralization and overall health (Jiajue et al., 2019). The widespread occurrence of low levels of Vitamin D in many regions of the world has garnered significant attention from researchers, medical professionals, and public health due to its connections with various illnesses. Deficiency in Vitamin D can lead to loss of bone density, which can contribute to rickets in children and osteomalacia and osteoporosis in adults (Jiajue et al., 2019).

For a long period of time, it was thought that the only role of Vitamin D was in maintaining healthy bones. Late research suggests that insufficient Vitamin D levels may contribute to the development of certain types of cancer, such as colon, breast and prostate cancer (K. Amrein et al., 2020). Low levels of Vitamin D have also been linked to an increased risk of diabetes, cardiovascular disease, autoimmune diseases**,** such as multiple sclerosis and rheumatoid arthritis, and infections, such as tuberculosis (K. Amrein et al., 2020). Further investigations are required to gain a more comprehensive understanding of the impact of Vitamin D at the molecular level on health and illnesses beyond its traditional role in regulating calcium levels.

Recent technological advancements, specifically the use of next-generation deoxyribonucleic acid (DNA) sequencing in functional genomics **(Hypponen, Vimaleswaran, & Zhou, 2022)**, allow for high-scale studies that aim to address questions related to DNA-protein interactions with more precision and detail than ever before. In this chapter, I will examine the existing research on the sources of Vitamin D, its metabolism, functions, and deficiency. I will focus on work that provides insight into the genomic associations with Vitamin D.

1.1.1 *History and Evolutionary Perspective of Vitamin D*

The research on childhood illness rickets resulted in identifying the precursor secosteriod hormone known as Vitamin D. Rickets is a pediatric orthopedic disorder characterized by stunted growth, skeletal deformities, bony protuberances on the rib cage (sometimes referred to as a "rachitic rosary"), and either bowed or knock-kneed legs due to decreased skeletal mineralization (M. F. Holick, 2004). The disease was first described by the English physician Francis Glisson in 1650 (Rajakumar, Greenspan, Thomas, & Holick, 2007). Rickets became a widespread rampant disease among European children during the Industrial Revolution (M. F. Holick, 2004; Palm, 1890). In the early 19^{th} century, healthcare providers began successfully preventing and treating children from rickets with liver oil from fish, and later by exposing them to sunlight or lamps emitting mercury vapor (Eliot, 1925; Huldschinsky, 1919). However, the specific cause of rickets, a deficiency of Vitamin D, was discovered in the early 20^{th} century (Rajakumar et al., 2007).

The detection was made by a number of different clinicians, including Edward Mellanby, who showed in 1919 that rickets could be caused by a lack of Vitamin D (Mellanby, 1919), and later on, the discovery of the active form of Vitamin D was made by Adolf Windaus in 1922 (Wolf, 2004). The discovery that Vitamin D can be synthesized through exposure to sunlight led to significant improvements in the management of rickets. The implementation of fortification programs with Vitamin D was the first successful strategy in decreasing the prevalence of rickets, with the United States saw a nearly complete eradication of the disease by the 1960s. Currently, it is mandatory to fortify margarine and infant formula with Vitamin D in the United Kingdom (Rajakumar et al., 2007).

From an evolutionary perspective, it is believed that the ability to synthesize Vitamin D from sunlight has been an essential adaptation for many species. Some studies suggest that organisms older than 500 million years old, such as phytoplankton (*Emiliania huxleyi*) and diatoms (*Skeletonema menxelii*), have the ability to synthesize Vitamin D from ergosterol or Vitamin D2 (Michael F Holick, Pang, & Schreibman, 1989). However, the exact way Vitamin D evolved and functions in non-vertebrates still need to be fully understood. Conversely, Vitamin D is crucial in regulating the intracellular and extracellular calcium and phosphorous levels in vertebrates, which helps develop skeletal and other metabolic functions properly (Cutie, Payumo, Lunn, & Huang, 2020).

It is hypothesized that Vitamin D and its precursors in humans and animals of the ancient time have had a role in the mechanism of protection against ultraviolet (UV) radiation upon exposure to sunlight (Ames, Grant, & Willett, 2021; Michael F Holick et al., 1989). However, this adaptation decreases the capacity of Vitamin D synthesis from sunlight.

Noteworthy, Vitamin D deficiency is more common in individuals living in regions with less sunlight exposure, such as in high latitudes, darker skin pigmentation, elderly, and overweight or obese people (Ames et al., 2021). This phenomenon is believed to be an evolutionary adaptation to these individuals.

Additionally, studies in molecular biology have shown that the genetic machinery necessary for Vitamin D synthesis is highly conserved across many different species, indicating its long-standing biological importance (Azarpeykan et al., 2016; Girgis et al., 2019). Overall, Vitamin D synthesis through sunlight exposure is a critical adaptation that has allowed organisms to survive and thrive in various environments throughout evolutionary history.

1.1.2 *Sources of Vitamin D*

Vitamin D is produced in the skin when exposed to UVB light from the sun at wavelengths of 290 to 315 nanometer (K. Amrein et al., 2020). However, the amount of solar UVB radiation that is necessary to produce the appropriate amount of Vitamin D can vary depending on factors, such as time of day, season, and latitude. In such cases of limited sunlight, the most effective way is to obtain an adequate level of Vitamin D through dietary sources. The Institute of Medicine in the United States suggests a daily intake of 600-800 International Units (IU) of Vitamin D for adults, while the Endocrine Society recommends a higher dose of 1500-2000 IU per day (K. Amrein et al., 2020).

The primary sources of exogenous Vitamin D are Vitamin D2 (ergocalciferol) and Vitamin D3 (cholecalciferol). Vitamin D2 can be found in plant-based sources such as mushrooms, fortified bread and cereals, and certain types of fish. Sun-dried mushrooms are an excellent plant source, providing around 1600 IU of Vitamin D2 per 3.5 ounces (Duffy et al., 2018). Vitamin D3 can be found in animal-based foods, such as egg yolks, fatty fish, and fish oils. Among the best natural sources of Vitamin D3 are fatty fish, such as fresh wild salmon, which provides 600 to 1000 IU per 3.5 oz serving, and fish oils, such as cod liver oil, which provides 400 to 1000 IU per 1 tablespoon (Duffy et al., 2018).

The two forms are similar but have a slight structural difference. Vitamin D2 has a methyl group at carbon 24 and a double bond between carbon 22 and 23, which makes it less effective than D3 (Heaney, Recker, Grote, Horst, & Armas, 2011; Shieh et al., 2016). Due to this, Vitamin D3 has become more preferred for use in supplements and fortified foods to treat and prevent deficiency. However, there is no differentiation observed between Vitamin D2 and D3 in the literature, and both forms are referred to simply as "Vitamin D."

1.1.3 *Synthesis and Metabolism of Vitamin D*

1.1.3.1 *Cutaneous Synthesis of Vitamin D*

Vitamin D is produced in the skin of humans and animals through a series of chemical reactions that involve different proteins (Figure 1). The first step is converting a cholesterol derivative, called 7-dehydrocholesterol (7-DHC), abundant in lipid membranes of skin cells, through a UVB photoisomerization to pre-Vitamin D3 (Prabhu, Luu, Sharpe, & Brown, 2016).

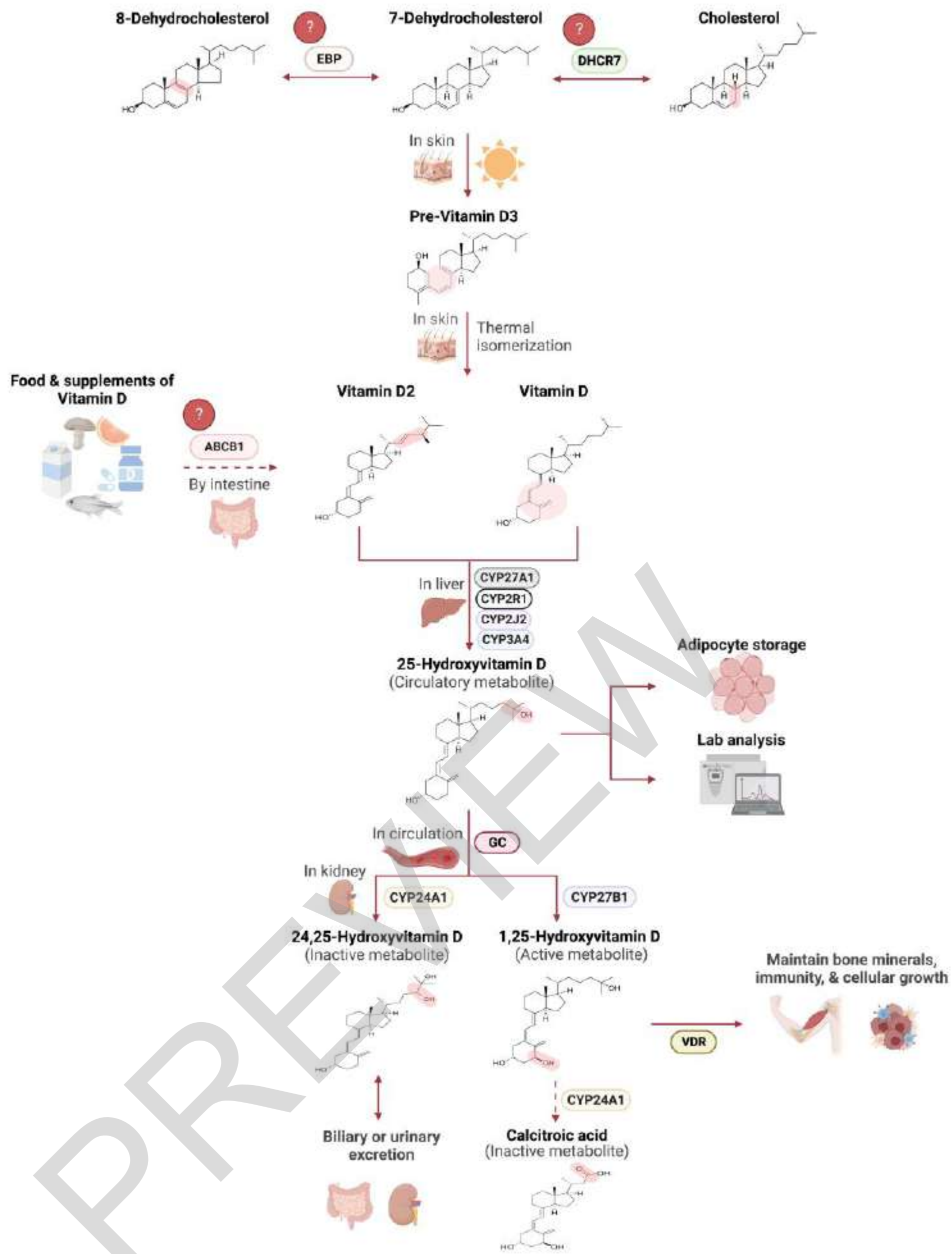

Figure 1 Vitamin D Synthesis and Metabolism.

The pathway starts with the cutaneous synthesis of Vitamin D3 from 7-dehydrocholesterol upon solar ultraviolet B (UVB) exposure. 7-dehydrocholesterol can also be synthesized from 8-dehydrocholesterol or cholesterol by EBP or DHCR7, respectively. Vitamin D can also be absorbed through intestinal food by ABCB1. The conversion to 25-hydroxyVitamin D, commonly used for Vitamin D's status analysis, occurs in the liver by CYP27A1, CYP2R1, CYP2J2, or CYP3A4 and then activated to 1,25-dihydroxyVitamin D by renal CYP27B1 or stored in the body for later use. The active form regulates gene expression through the nuclear Vitamin D Receptor (VDR)/ Retinoid-X Receptor (RXR) for overall health before being inactivated by several enzymes, e.g., CYP24A1, and excreted. (?) Indicates other enzymes involved in Vitamin D metabolism but has not been discovered yet. 2D chemical structures obtained from PubChem: https://pubchem.ncbi.nlm.nih.gov. Abbreviation: DHCR7, 7-Dehydrocholesterol reductase; ABCB1, ATP-binding cassette transporter B1; CYP, Cytochrome P450. For a complete listing of the SNPs associated with each gene, please refer to Table 2. Generated with BioRender.com.