# A Pathway Enrichment Approach in deciphering the Genetic Basis of Vitamin D Insufficiency

**Sravana Deepthi C.[1], Ravikanth Eddula Venkata[2]*, Adiga Usha[3], Vasishta Sampara[4] and Sindhu B.S.[2]**
1. Department of Community Medicine, Apollo Institute of Medical Sciences and Research Chittoor, Andhra Pradesh, INDIA
2. Department of Dermatology, Apollo Institute of Medical Sciences and Research Chittoor, Andhra Pradesh, INDIA
3. Department of Biochemistry, Apollo Institute of Medical Sciences and Research Chittoor, Andhra Pradesh, INDIA
4. ICMR, Department of Biochemistry, Apollo Institute of Medical Sciences and Research Chittoor, Andhra Pradesh, INDIA
*ravikanthmikado@gmail.com

## *Abstract*

*A number of skeletal and non-skeletal conditions are linked to vitamin D deficiency, which is a major public health concern. Numerous genetic loci connected to the metabolism and regulation of vitamin D have been found by genome-wide association studies (GWAS). In order to identify possible indicators and treatment targets, this project will examine GWAS data to identify genetic variations causing vitamin D insufficiency. To find significant single nucleotide polymorphisms (SNPs) linked to vitamin D levels, this study used data obtained from the GWAS collection. These genetic variations were mapped to biological processes, such as immune regulation, metabolism and vitamin D production, using enrichment analysis. To evaluate their functional implications, interactions between the discovered genes and regulatory elements including transcription factors and miRNAs were also investigated. Vitamin D deficiency was substantially linked to important genetic loci, such as NADSYN1. Pathways pertaining to immunological modulation, mitochondrial activity and lipid metabolism were identified using enrichment analysis. Numerous miRNAs have been found to be putative regulators of genes related to the synthesis and metabolism of vitamin D. These results point to new molecular targets for more research and point to a hereditary basis for vitamin D deficiency. In order to help with risk stratification, tailored prevention and therapeutic targeting for better public health interventions, this work uses a secondary bioinformatics analysis of GWAS data from Wang et al[16] to find genetic variations associated with vitamin D deficiency.*

**Keywords:** Vitamin D insufficiency, GWAS, genetic variants, enrichment analysis, metabolic pathways**.**

## Introduction

Numerous physiological functions such as calcium and phosphate balance, bone mineralization, immunological regulation and cardiovascular health, depend heavily on vitamin D.[14] Nonetheless, vitamin D deficiency continues to be a major worldwide health concern, raising the risk of viral diseases, metabolic syndromes, autoimmune diseases and osteoporosis.[1,5] By finding important genetic factors impacting blood 25-hydroxyvitamin D [25(OH)D] levels, genome-wide association studies (GWAS) have become a potent method for elucidating the genetic basis of vitamin D deficiency.[12,16]

Numerous genetic variations within the vitamin D metabolic route have been identified by recent GWAS, including those encoding the vitamin D receptor (VDR), cytochrome P450 enzymes (CYP2R1, CYP24A1) and vitamin D-binding protein (GC).[3,6] Individual disparities in vitamin D status across various groups are caused by these genetic polymorphisms, which also affect vitamin D production, transport and breakdown.[2,13] Identification of individuals at risk, development of focused dietary therapies and improvement of personalized medicine methods can all be facilitated by an understanding of these genetic foundations.[9]

Based on the research of Wang et al[16] who found common genetic drivers of vitamin D insufficiency, analysis makes use of data from the GWAS database. This work investigates the functional relevance of genes linked to serum vitamin D levels by mapping them, offering insights on hereditary vulnerability and possible targets for enhancing vitamin D status. By combining gene ontology (GO) analysis, pathway enrichment, post-transcriptional regulation, protein-protein interactions and metabolomic linkages, this study intends to use GWAS data to analyze mapped genes linked to vitamin D insufficiency and conduct a thorough multi-omics approach.

This work attempts to offer new insights into the genetic and molecular mechanisms driving vitamin D insufficiency by combining GWAS data with functional and pathway studies. Better risk categorization, enhanced diagnostic biomarkers and possible treatment targets for reducing vitamin D-related health effects will benefit all.

## Material and Methods

**Ethical consideration:** This study exclusively utilizes publicly available data and does not involve human participants, animal subjects, or any identifiable personal information. As a result, institutional ethics committee approval was not required. Nonetheless, the research adheres to established ethical standards including the Declaration of Helsinki and relevant regulatory guidelines. All data sources comply with open-access policies and proper citations have been included to credit the original data providers.

**Inclusion criteria:** This study included publicly available datasets from the GWAS catalogue focused on mapped

genes, specifically those involving vitamin D insufficiency. The datasets providing comprehensive metadata and raw or processed data suitable for differential gene expression analysis were selected.

**Exclusion criteria:** Datasets unrelated to vitamin D insufficiency, those with incomplete data hindering analysis, or those failing quality control for normalization and reliability were excluded. We used a multi-step bioinformatics strategy that included genome-wide association study (GWAS) data, functional enrichment analysis, pathway mapping and network-based investigations to look into the genetic and molecular pathways underlying vitamin D insufficiency.

Initially, information is obtained from the GWAS catalogue in order to determine which genes are substantially linked to vitamin D deficiency. These mapped genes were taken out and selected for additional analysis later on. [16] The reliability of the chosen genetic variations was ensured by applying statistical thresholds and quality control procedures, with an emphasis on those that have been shown to be relevant to the metabolism and regulation of vitamin D. The mapped genes were then categorized according to their biological roles using gene ontology (GO) analysis. In order to provide insight into how these genes contribute to vitamin D homeostasis, it included classifying genes based on biological processes, molecular activities and cellular components.

Key functional categories and molecular pathways including calcium homeostasis, bone mineralization and endocrine control, associated with vitamin D deficiency were highlighted using enrichment analysis utilizing well-established techniques. The KEGG and Reactome databases were used to do pathway analysis in order to further break down disease processes. In this phase, we were able to ascertain whether the genes that were found were abundant in pathways related to vitamin D, including calcium absorption, vitamin D metabolism, VDR (Vitamin D Receptor) signaling, PTH (parathyroid hormone) regulation and immune system modulation. The findings provided a systems level understanding of how genetic differences affect the synthesis, metabolism and function of vitamin D.

Using miRNA enrichment analysis, post-transcriptional and epigenetic regulation mechanisms were investigated. We found miRNAs that target important genes involved in vitamin D metabolism and signaling by using the miRTarBase and TargetScan databases. This investigation shed light on the regulatory networks that affect vitamin D-mediated immune responses and gene expression, both of which may raise the risk of insufficiency. Determining possible targets for dietary and medicinal interventions requires an understanding of these interconnections.

The functional connection between genes linked to vitamin D deficiency was investigated using a protein-protein interaction (PPI) network analysis. By using this method, we were able to identify genes that have strong functional relationships and may be important for maintaining vitamin D homeostasis and the pathophysiology associated with its insufficiency.

Finally, by using MetaboAnalyst to examine whether mapped genes impact metabolic pathways related to vitamin D shortage, we connected GWAS results to metabolomic alterations. Potential metabolic indicators linked to calcium homeostasis, vitamin D status and bone health were found with the aid of this investigation. Our goal was to find new biomarkers that could enhance the prediction of vitamin D deficiency and aid in early diagnosis by combining genetic and metabolomic data.

## Results

Table 1 highlights important loci on chromosomes 4 and 11 and displays genomic areas linked to vitamin D metabolism. NADSYN1, DHCR7, GC and CYP2R1 are among the genes that have been identified, these genes are essential for the manufacture of cholesterol, the transport of vitamin D and metabolic activation. These correlations are supported by the mapped genes. CYP2R1 and CALCB were found together, indicating possible new interactions. These results underline the importance of genetic control over vitamin D levels and the necessity for more investigation into the potential functional effects of these genetic variations on vitamin D metabolism.

Key genes found in many tissues and animals are highlighted in table 2 which displays cell marker connections for 2024. In mice with a high combined score, CALCB is associated with retinal cells, suggesting great statistical significance. Krt4/13+ tracheal cells are also linked to it, albeit with a lower chances ratio. The identification of NADSYN1 in the fetal human gonad's Leydig cells raises the possibility of developmental function. The strongest correlation was found in retinal cells and the adjusted p-values show statistical robustness. These results imply that CALCB and NADSYN1 may have functional roles, in particular, cellular contexts.

TargetScan 2017 microRNA interactions are shown in table 3, suggesting possible regulatory relationships between NADSYN1 and CALCB and other miRNAs. Several human miRNAs, such as hsa-miR-3195, hsa-miR-4787-3p, hsa-miR-3131 and hsa-miR-611, have high combined scores that indicate strong binding potential and target NADSYN1. Significant enrichment is suggested by the odds ratios, even with comparatively large p-values.

These interactions suggest that NADSYN1 and CALCB may be post-transcriptionally regulated, impacting their biological roles in cellular processes. The effect of these miRNAs on gene expression and disease processes may become clearer with more research. Based on GWAS data, the table 4 displays microRNAs (miRNAs) connected to

NADSYN1, a gene involved in vitamin D metabolism. Indicating statistical importance, it contains miRNA names, overlap counts, p-values, corrected p-values, odds ratios and combined scores. The highest total score is displayed by hsa-miR-4320, indicating substantial regulatory potential. The significance of NADSYN1 in vitamin D deficiency is further supported by the notable enrichment of other miRNAs, including hsa-miR-5004-3p and hsa-miR-4503. The results provide information on genetic vulnerability and prospective treatment strategies for vitamin D deficiency by highlighting putative post-transcriptional regulatory mechanisms influencing vitamin D biosynthesis and metabolism.

Based on reactome pathways 2024 research, table 5 lists important biological pathways linked to the genes CYP2R1, GC, NADSYN1 and CALCB. With the lowest p-value (1.98E-06) and the highest combined score (26247.83), the most enriched pathway, vitamin D (Calciferol) metabolism, highlights the critical roles that CYP2R1 and GC play in the metabolism of vitamin D. Other noteworthy routes that connect CYP2R1 and NADSYN1 to the metabolism of steroids and vitamins include metabolism of steroids and nicotinate metabolism. The strong correlation between CALCB and calcitonin-like ligand receptors suggests that it plays a part in hormone signaling. The function of CYP2R1 in enzymatic oxidation is highlighted by the cytochrome P450 pathway. These findings imply that the discovered genes have a significant metabolic and hormonal regulatory role which calls for more research.

The KEGG 2021 human database's enriched pathways linked to CYP2R1, NADSYN1 and CALCB are shown in table 6. The most highly enriched route is steroid biosynthesis ($p = 0.003994$, total score = 1935.63), underscoring CYP2R1's function in steroid metabolism. Another substantially enriched pathway ($p = 0.006982$) that connects NADSYN1 to processes linked to vitamin B3, is the metabolism of nicotinate and nicotinamide. Vascular smooth muscle contraction ($p = 0.026338$) and neuroactive ligand-receptor interaction ($p = 0.06648$) are linked to CALCB, indicating a potential involvement for the protein in neurological and cardiovascular processes.

**Table 1**
**GWAS findings of Vitamin D insufficiency**

| Region | CHR_ID | CHR_POS | Reported Gene (S) | Mapped_Gene |
|---|---|---|---|---|
| 11q13.4 | 11 | 71456403 | NADSYN1, DHCR7 | NADSYN1 |
| 4q13.3 | 4 | 71742666 | GC | GC |
| 11p15.2 | 11 | 14893332 | CYP2R1 | CYP2R1 - CALCB |

**Table 2**
**Cell marker 2024**

| Term | Overlap | P-value | Adjusted P-value | Old P-value | Old Adjusted P-value | Odds Ratio | Combined Score | Genes |
|---|---|---|---|---|---|---|---|---|
| Leydig Cell Fetal Gonad Human | 1/383 | 0.074433 | 0.074433 | 0 | 0 | 17.11518 | 44.46283342 | NADSYN1 |

**Table 3**
**TargetScan_microRNA_2017**

| Term | Overlap | P-value | Adjusted P-value | Old P-value | Old Adjusted P-value | Odds Ratio | Combined Score | Genes |
|---|---|---|---|---|---|---|---|---|
| hsa-miR-3195 | 1/298 | 0.058285 | 0.343776 | 0 | 0 | 22.10887 | 62.84241926 | NADSYN1 |
| hsa-miR-4787-3p | 1/365 | 0.071031 | 0.343776 | 0 | 0 | 17.97802 | 47.54544256 | NADSYN1 |
| hsa-miR-3131 | 1/569 | 0.109043 | 0.343776 | 0 | 0 | 11.40141 | 25.26571652 | NADSYN1 |
| hsa-miR-611 | 1/569 | 0.109043 | 0.343776 | 0 | 0 | 11.40141 | 25.26571652 | NADSYN1 |
| hsa-miR-3615 | 1/774 | 0.146053 | 0.343776 | 0 | 0 | 8.289349 | 15.94690718 | NADSYN1 |
| hsa-miR-1307 | 1/974 | 0.181038 | 0.343776 | 0 | 0 | 6.516958 | 11.13778025 | NADSYN1 |
| hsa-miR-4638-5p | 1/1196 | 0.218601 | 0.343776 | 0 | 0 | 5.244351 | 7.974077078 | NADSYN1 |

**Table 4**
**miRTarBase_2017**

| Term | Overlap | P-value | Adjusted P-value | Old P-value | Old Adjusted P-value | Odds Ratio | Combined Score | Genes |
|---|---|---|---|---|---|---|---|---|
| hsa-miR-4320 | Jan-49 | 0.009765 | 0.040437 | 0 | 0 | 138.5278 | 641.2429069 | NADSYN1 |
| hsa-miR-5004-3p | Jan-54 | 0.010757 | 0.040437 | 0 | 0 | 125.4277 | 568.462712 | NADSYN1 |
| hsa-miR-4503 | Jan-71 | 0.014125 | 0.040437 | 0 | 0 | 94.88571 | 404.1917114 | NADSYN1 |
| hsa-miR-2682-3p | 1/101 | 0.020049 | 0.040437 | 0 | 0 | 66.32 | 259.2836075 | NADSYN1 |
| hsa-miR-33b-5p | 1/101 | 0.020049 | 0.040437 | 0 | 0 | 66.32 | 259.2836075 | NADSYN1 |
| hsa-miR-6781-3p | 1/101 | 0.020049 | 0.040437 | 0 | 0 | 66.32 | 259.2836075 | NADSYN1 |
| hsa-miR-6877-3p | 1/117 | 0.023197 | 0.040437 | 0 | 0 | 57.12644 | 215.008552 | NADSYN1 |

**Table 5**
**Reactome_Pathways_2024**

| Term | Overlap | P-value | Adjusted P-value | Old P-value | Old Adjusted P-value | Odds Ratio | Combined Score | Genes |
|---|---|---|---|---|---|---|---|---|
| Vitamin D (Calciferol) Metabolism | 02-Dec | 1.98 E-06 | 4.16E-05 | 0 | 0 | 1998. 6 | 26247. 83934 | CYP2R1;GC |
| Metabolism of Steroids | 2/155 | 3.54 E-04 | 0.003721 | 0 | 0 | 129. 6928 | 1030. 414873 | CYP2R1;GC |
| Vitamins | 01-Jun | 0.0012 | 0.008397 | 0 | 0 | 1332. 733 | 8963. 738321 | CYP2R1 |
| Calcitonin-like Ligand Receptors | 01-Oct | 0. 001999 | 0.010493 | 0 | 0 | 740. 2593 | 4600. 934596 | CALCB |
| Metabolism | Mar-81 | 0. 004758 | 0.019982 | 0 | 0 | 24. 5427 | 131. 2545984 | CYP2R1;NADSYN1;GC |
| Nicotinate Metabolism | Jan-31 | 0. 006186 | 0.020349 | 0 | 0 | 221. 8444 | 1128. 183024 | NADSYN1 |
| Metabolic Disorders of Biological Oxidation Enzymes | Jan-34 | 0. 006783 | 0.020349 | 0 | 0 | 201. 6465 | 1006. 885371 | CYP2R1 |
| Metabolism of Lipids | 2/758 | 0. 00818 | 0.021472 | 0 | 0 | 25. 44974 | 122. 3139539 | CYP2R1;GC |
| Cytochrome P450 - Arranged by Substrate Type | Jan-65 | 0. 012938 | 0.030188 | 0 | 0 | 103. 8125 | 451. 3369693 | CYP2R1 |
| Class B 2 (Secretin Family Receptors) | Jan-94 | 0. 018669 | 0.039205 | 0 | 0 | 71. 33692 | 283. 9840279 | CALCB |

According to these results, the discovered genes support important physiological functions such as metabolism, vascular control and neurotransmission, which call for more research into their precise biological functions.

The enriched gene ontology (GO) biological processes (2025) linked to CYP2R1, GC, NADSYN1 and CALCB are shown in table 7. The vitamin D metabolic process (GO:0042359) is the most important process, emphasizing the roles of CYP2R1 and GC in vitamin D metabolism ($p = 1.98E-06$, combined score = 26247.84). The metabolisms of fat-soluble vitamins (GO:0006775) and steroids (GO:0008202) are additional important processes that highlight the function of CYP2R1 in lipid-soluble nutrients. NAD biosynthesis (GO:0009435, $p = 0.001599$), which is essential for cellular energy metabolism, is associated with

NADSYN1. An involvement in hormone signaling is suggested by CALCB's association with the calcitonin receptor signaling pathway (GO:0097646, p = 0.001799). These results highlight the various metabolic and signaling roles that these genes play.

The gene ontology (GO) cellular components (2025) linked to GC and CYP2R1 are highlighted in table 8. Lysosomal Lumen (GO:0043202, p = 0.017091, combined score = 317.73) is the most enriched word, suggesting that GC is involved in lysosomal function. Furthermore, GC's involvement in intracellular breakdown pathways is reinforced by its association with the vacuolar lumen (GO:0005775, p = 0.032206). These results are further supported by the lysosome relationship (GO:0005764, p = 0.102236). CYP2R1's known role in vitamin D metabolism is consistent with its primary association with the endoplasmic reticulum membrane (GO:0005789, p = 0.163333). A wider distribution of CYP2R1 across cellular compartments is suggested by the reduced importance of the intracellular membrane-bounded organelle (GO:0043231, p = 0.731068).

The gene ontology (GO) molecular functions (2025) linked to CYP2R1, CALCB, GC and NADSYN1 are shown in table 9. The most important function highlights the significance of NADSYN1 in NAD biosynthesis: carbon-nitrogen ligase activity with glutamine as an amido-N donor (GO:0016884, p = 0.001399, combined score = 7298.29).

**Table 6**
**KEGG_2021_Human**

| Term | Overlap | P-value | Adjusted P-value | Old P-value | Old Adjusted P-value | Odds Ratio | Combined Score | Genes |
|---|---|---|---|---|---|---|---|---|
| Steroid biosynthesis | Jan-20 | 0.003994 | 0.013964 | 0 | 0 | 350.4737 | 1935.631384 | CYP2R1 |
| Nicotinate and nicotinamide metabolism | Jan-35 | 0.006982 | 0.013964 | 0 | 0 | 195.7059 | 971.5637664 | NADSYN1 |
| Vascular smooth muscle contraction | 1/133 | 0.026338 | 0.035117 | 0 | 0 | 50.16162 | 182.4256037 | CALCB |
| Neuroactive ligand-receptor interaction | 1/341 | 0.06648 | 0.06648 | 0 | 0 | 19.27059 | 52.23971901 | CALCB |

**Table 7**
**GO_Biological_Process_2025**

| Term | Overlap | P-value | Adjusted P-value | Old P-value | Old Adjusted P-value | Odds Ratio | Combined Score | Genes |
|---|---|---|---|---|---|---|---|---|
| Vitamin D Metabolic Process (GO:0042359) | 02-Dec | 1.98 E-06 | 3.56 E-05 | 0 | 0 | 1998.6 | 26247.83934 | CYP2R1;GC |
| Fat-Soluble Vitamin Metabolic Process (GO:0006775) | Feb-26 | 9.73 E-06 | 8.76 E-05 | 0 | 0 | 832.1667 | 9603.084843 | CYP2R1;GC |
| Steroid Metabolic Process (GO:0008202) | 2/106 | 1.66 E-04 | 9.95 E-04 | 0 | 0 | 191.2692 | 1664.950312 | CYP2R1;GC |
| Vitamin Metabolic Process (GO:0006766) | 01-May | 1.00 E-03 | 0.004047 | 0 | 0 | 1666 | 11508.8581 | CYP2R1 |
| Nicotinamide Nucleotide Biosynthetic Process (GO:0019359) | 01-Aug | 0.001599 | 0.004047 | 0 | 0 | 951.8571 | 6128.338527 | NADSYN1 |
| NAD Biosynthetic Process (GO:0009435) | 01-Aug | 0.001599 | 0.004047 | 0 | 0 | 951.8571 | 6128.338527 | NADSYN1 |
| Calcitonin Family Receptor Signaling Pathway (GO:0097646) | 01-Sep | 0.001799 | 0.004047 | 0 | 0 | 832.8333 | 5263.996189 | CALCB |
| Organic Acid Metabolic Process (GO:0006082) | 01-Sep | 0.001799 | 0.004047 | 0 | 0 | 832.8333 | 5263.996189 | CYP2R1 |
| Polyol Biosynthetic Process (GO:0046173) | Jan-14 | 0.002797 | 0.005594 | 0 | 0 | 512.3846 | 3012.374505 | CYP2R1 |
| NAD Metabolic Process (GO:0019674) | Jan-18 | 0.003595 | 0.006472 | 0 | 0 | 391.7451 | 2204.785214 | NADSYN1 |

CYP2R1's role in steroid metabolism is supported by its association with iron ion binding (GO:0005506, p = 0.011352) and steroid hydroxylase activity (GO:0008395, p = 0.004592). CALCB may have a role in signaling because it is implicated in both neuropeptide hormone activity (GO:0005184, p = 0.00519) and hormone activity (GO:0005179, p = 0.017485). Furthermore, actin binding (GO:0003779, p = 0.036881) is linked to GC and may be involved in cytoskeletal interactions.

The human metabolome database (HMDB) compounds linked to CYP2R1 and NADSYN1 are shown in table 10. NADSYN1 is strongly associated with L-glutamine (HMDB00641, p = 0.005389, combined score = 1337.30) and L-glutamic acid (HMDB00148, p = 0.01155), highlighting its function in NAD production and amino acid metabolism. NADSYN1's function in cellular energy metabolism is further supported by its interactions with NAD (HMDB00902, p = 0.048303), NADH (HMDB01487, p = 0.040183) and NADPH (HMDB00221, p = 0.034351). CYP2R1 may have a role in oxidative processes as it is associated with C34H34N4O4.Fe (HMDB03178, p = 0.033376) and NAP (HMDB00217, p = 0.033571). Furthermore, adenosine monophosphate (HMDB00045, p = 0.046181) indicates a role for nucleotide metabolism.

The protein-protein interaction (PPI) hub proteins linked to the GC gene are highlighted in table 11, indicating that it may be involved in a number of biological functions. The strongest correlation is shown by C1ORF103 (p = 0.024179, combined score = 203.80), which may connect GC to unidentified cellular processes.

Given its function in DNA replication and repair, RIF1 (p = 0.031034, combined score = 147.22) may be linked to genomic stability and GC. GC may have a structural or signaling role, according to ACTA1 (p = 0.032596, combined score = 138.00), a crucial actin cytoskeleton component. The growth factor signaling protein GRB2 (p = 0.144809, combined score = 16.17) exhibits a modest but significant correlation, suggesting that GC may play a part in cell signaling pathways.

Table 12 lists downregulated gene signatures from RNA sequencing (RNA-seq) investigations that show correlations with CALCB, NADSYN1 and CYP2R1 under different circumstances. Potential functions in inflammatory and metabolic pathways are suggested by the presence of CYP2R1 in datasets such as peripheral cervical lesion analysis (GSE120691) and IL-1B stimulation in chondrocytes (GSE74220). NADSYN1 is implicated in gene regulation and cellular signaling, as evidenced by its association with Erbb2/Igf1R signaling (GSE73628) and mir-100/mir-125B regulation (GSE88757). Pre-mRNA haploinsufficiency modeling (GSE129619) shows CALCB, suggesting a potential role in neuroendocrine or RNA processing. Every term displays the same total score (79.69) and p-value (0.049074), suggesting that gene downregulation is statistically significant at a consistent level across investigations.

Upregulated gene profiles from RNA sequencing (RNA-seq) investigations are shown in table 13, emphasizing relationships with CYP2R1, GC, CALCB and NADSYN1. CALCB and GC have the strongest relationships (p = 9.18E-04, adjusted p = 0.045614) with hepatitis NS5A-related growth regulation (GSE102910), LOXL2 depletion in esophageal tissue (GSE135510) and GATA6 haploinsufficiency (GSE92581), indicating roles in immunological response, fibrosis and development. NADSYN1 is implicated in inflammation and metabolic control, as evidenced by its association with estrogen receptor-MDM2 inhibition (GSE140758) and generalized pustular psoriasis (GSE123785). CYP2R1 may have an impact on metabolic pathways and oxidative stress in keratoconus (GSE151631) and LOXL2 depletion (GSE135510). Potential molecular mechanisms underlying disease processes are revealed by these results

**Table 8**
**GO_Cellular_Component_2025**

| Term | Overlap | P-value | Adjusted P-value | Old P-value | Old Adjusted P-value | Odds Ratio | Combined Score | Genes |
|---|---|---|---|---|---|---|---|---|
| Lysosomal Lumen (GO:0043202) | Jan-86 | 0.017091 | 0.080515 | 0 | 0 | 78.08235 | 317.7351936 | GC |
| Vacuolar Lumen (GO:0005775) | 1/163 | 0.032206 | 0.080515 | 0 | 0 | 40.8107 | 140.20957 | GC |
| Lysosome (GO:0005764) | 1/532 | 0.102236 | 0.170394 | 0 | 0 | 12.21908 | 27.86521431 | GC |
| Endoplasmic Reticulum Membrane (GO:0005789) | 1/872 | 0.163333 | 0.204167 | 0 | 0 | 7.319173 | 13.26207007 | CYP2R1 |
| Intracellular Membrane-Bounded Organelle (GO:0043231) | Jan-97 | 0.731068 | 0.731068 | 0 | 0 | 0.857756 | 0.268690788 | CYP2R1 |

**Table 9**
**GO_Molecular_Function_2025**

| Term | Overlap | P-value | Adjusted P-value | Old P-value | Old Adjusted P-value | Odds Ratio | Combined Score | Genes |
|---|---|---|---|---|---|---|---|---|
| Carbon-Nitrogen Ligase Activity, With Glutamine as amido-N-donor (GO:0016884) | 01-Jul | 0.001399 | 0.016852 | 0 | 0 | 1110.556 | 7298.297606 | NADSYN1 |
| Steroid Hydroxylase Activity (GO:0008395) | Jan-23 | 0.004592 | 0.016852 | 0 | 0 | 302.6364 | 1629.201456 | CYP2R1 |
| Neuropeptide Hormone Activity (GO:0005184) | Jan-26 | 0.00519 | 0.016852 | 0 | 0 | 266.28 | 1400.895167 | CALCB |
| Neuropeptide Activity (GO:0160041) | Jan-27 | 0.005389 | 0.016852 | 0 | 0 | 256.0256 | 1337.303804 | CALCB |
| Oxidoreductase, Paired Donors, W Incorp or Reduce of O2, Reduced Flavin or Flavoprotein as One Donor and Incorp of O (GO:0016712) | Jan-36 | 0.007181 | 0.016852 | 0 | 0 | 190.1048 | 938.416342 | CYP2R1 |
| Hydrolase Activity, Acting on Carbon-Nitrogen (But Not Peptide) Bonds, in Linear Amides (GO:0016811) | Jan-39 | 0.007778 | 0.016852 | 0 | 0 | 175.0702 | 850.2271896 | NADSYN1 |
| Iron Ion Binding (GO:0005506) | Jan-57 | 0.011352 | 0.021082 | 0 | 0 | 118.6905 | 531.5378491 | CYP2R1 |
| Hormone Activity (GO:0005179) | Jan-88 | 0.017485 | 0.028414 | 0 | 0 | 76.27969 | 308.6575655 | CALCB |
| G Protein-Coupled Receptor Binding (GO:0001664) | 1/143 | 0.028297 | 0.040873 | 0 | 0 | 46.60563 | 166.1496472 | CALCB |
| Actin Binding (GO:0003779) | 1/187 | 0.036881 | 0.047946 | 0 | 0 | 35.50179 | 117.15776 | GC |

**Table 10**
**HMDB_Metabolites**

| Term | Overlap | P-value | Adjusted P-value | Old P-value | Old Adjusted P-value | Odds Ratio | Combined Score | Genes |
|---|---|---|---|---|---|---|---|---|
| L-Glutamine (HMDB00641) | Jan-27 | 0.005389 | 0.042351 | 0 | 0 | 256.0256 | 1337.303804 | NADSYN1 |
| Famotidine (HMDB01919) | Jan-44 | 0.008772 | 0.042351 | 0 | 0 | 154.6744 | 732.5752068 | NADSYN1 |
| L-Glutamic acid (HMDB00148) | Jan-58 | 0.01155 | 0.042351 | 0 | 0 | 116.6023 | 520.1672582 | NADSYN1 |
| C34H34N4O4.Fe (HMDB03178) | 1/169 | 0.033376 | 0.051383 | 0 | 0 | 39.34127 | 133.7567816 | CYP2R1 |
| NAP (HMDB00217) | 1/170 | 0.033571 | 0.051383 | 0 | 0 | 39.10651 | 132.7308345 | CYP2R1 |
| NADPH (HMDB00221) | 1/174 | 0.034351 | 0.051383 | 0 | 0 | 38.19461 | 128.7589409 | CYP2R1 |
| NADH (HMDB01487) | 1/204 | 0.040183 | 0.051383 | 0 | 0 | 32.50082 | 104.4679582 | NADSYN1 |
| Adenosine monophosphate (HMDB00045) | 1/235 | 0.046181 | 0.051383 | 0 | 0 | 28.151 | 86.56941049 | NADSYN1 |
| NAD (HMDB00902) | 1/246 | 0.048303 | 0.051383 | 0 | 0 | 26.87211 | 81.42951696 | NADSYN1 |
| Phosphoric acid (HMDB02142) | 1/261 | 0.051191 | 0.051383 | 0 | 0 | 25.30256 | 75.20427835 | NADSYN1 |

**Table 11**
**PPI_Hub_Proteins**

| Term | Overlap | P-value | Adjusted P-value | Old P-value | Old Adjusted P-value | Odds Ratio | Combined Score | Genes |
|---|---|---|---|---|---|---|---|---|
| C1ORF103 | 1/122 | 0.024179 | 0.043461 | 0 | 0 | 54.75207 | 203.8013894 | GC |
| RIF1 | 1/157 | 0.031034 | 0.043461 | 0 | 0 | 42.39316 | 147.2171233 | GC |
| ACTA1 | 1/165 | 0.032596 | 0.043461 | 0 | 0 | 40.30894 | 138.0002062 | GC |
| GRB2 | 1/767 | 0.144809 | 0.144809 | 0 | 0 | 8.368146 | 16.1701014 | GC |

**Table 12**
**RNAseq_Automatic_GEO_Signatures_Human_Down**

| Term | Overlap | P-value | Adjusted P-value | Old P-value | Old Adjusted P-value | Odds Ratio | Combined Score | Genes |
|---|---|---|---|---|---|---|---|---|
| Transcriptomics Fam122A Nc K562 GSE141735 1 | 1/250 | 0.049074 | 0.049074 | 0 | 0 | 26.43507 | 79.68675066 | CYP2R1 |
| Mir-100 Mir-125B Rip-Seq Ectopic GSE88757 2 | 1/250 | 0.049074 | 0.049074 | 0 | 0 | 26.43507 | 79.68675066 | NADSYN1 |
| Mir-100 Mir-125B Rip-Seq Ectopic GSE88757 4 | 1/250 | 0.049074 | 0.049074 | 0 | 0 | 26.43507 | 79.68675066 | NADSYN1 |
| 4H Il-1B Stimulation Chondrocytes GSE74220 1 | 1/250 | 0.049074 | 0.049074 | 0 | 0 | 26.43507 | 79.68675066 | CYP2R1 |
| Erbb2 Igf1R Raf1 Samples GSE73628 3 | 1/250 | 0.049074 | 0.049074 | 0 | 0 | 26.43507 | 79.68675066 | NADSYN1 |
| Cxcr5 Cd45Ra-Cd8 Subset Follicular GSE105095 1 | 1/250 | 0.049074 | 0.049074 | 0 | 0 | 26.43507 | 79.68675066 | CYP2R1 |
| Erbb2 Igf1R Raf1 Samples GSE73628 6 | 1/250 | 0.049074 | 0.049074 | 0 | 0 | 26.43507 | 79.68675066 | NADSYN1 |
| Differential Peripheral Cervical Lesions GSE120691 1 | 1/250 | 0.049074 | 0.049074 | 0 | 0 | 26.43507 | 79.68675066 | CYP2R1 |
| Modelling Core Pre-Mrna Haplo insufficiency GSE129619 1 | 1/250 | 0.049074 | 0.049074 | 0 | 0 | 26.43507 | 79.68675066 | CALCB |
| Study Role Pou3F2 Uvb GSE124761 1 | 1/250 | 0.049074 | 0.049074 | 0 | 0 | 26.43507 | 79.68675066 | NADSYN1 |

Figure 1 shows the ranking of the top 25 enriched metabolic pathways by enrichment ratio, which was determined using enrichment analysis. The color gradient indicates statistical significance (p-value), with darker red denoting greater importance. The urea cycle, glutamate metabolism and the metabolism of nicotinate and nicotinamide are important processes. Numerous metabolic processes, including lipid biosynthesis, energy production and amino acid metabolism, depend on these pathways. By connecting genetic differences to metabolic dysregulation, the enrichment analysis sheds light on the functional implications of these variations. Potential molecular pathways underlying the problem under study are shown in this graphic, providing useful targets for more study and treatment approaches.

## Discussion

Our genome-wide association study (GWAS) confirmed earlier findings on the genetic regulation of vitamin D metabolism by identifying important genetic loci linked to vitamin D deficiency. Important roles in vitamin D transport, activation and metabolic pathways are played by the discovered loci, which include GC, CYP2R1, NADSYN1 and CALCB. Notably, prior GWAS analyses have validated the extensive research on GC and CYP2R1's effects on serum 25-hydroxyvitamin D levels.[6,16]

**Gene Ontology (GO) Analysis:** Mapped genes were categorized by GO analysis according to cellular components, molecular activities and biological processes. The metabolism of vitamin D was the most enriched biological process, confirming the function of GC and CYP2R1 in the hydroxylation and transport of vitamin D.[12] Furthermore, the role of NADSYN1 in nicotinamide metabolism raises the possibility of a connection between energy metabolism and vitamin D homeostasis.[9]

**Pathway Analysis:** Numerous pathways related to vitamin D deficiency were found using KEGG and reactome

pathway enrichment analysis. The crucial functions of CYP2R1 and GC in vitamin D hydroxylation and transport were validated by the most important pathway, vitamin D metabolism.[3] Additionally, enhanced pathways like nicotinamide metabolism and steroid biosynthesis point to these genes' wider metabolic consequences.[4]

**Post-Transcriptional and Epigenetic Regulation:** Regulatory miRNAs that target genes linked to vitamin D were identified using miRNA enrichment analysis using miRTarBase and TargetScan. Several miRNAs, including hsa-miR-3195 and hsa-miR-4787-3p, were discovered to target NADSYN1, indicating post-transcriptional regulation that could affect vitamin D metabolism.[15]

**Protein-Protein Interaction (PPI) Network Analysis:** Key genes linked to vitamin D inadequacy were found to interact, according to PPI network analysis. Strong connections between GC, a key protein that binds vitamin D and structural and signaling proteins like ACTA1 and GRB2 suggested that GC may play a part in growth factor signaling and cytoskeletal dynamics.[11] These results point to new regulatory pathways in the homeostasis of vitamin D.

**Linking GWAS Findings to Metabolomic Changes:** The use of MetaboAnalyst for metabolomic analysis revealed that mapped genes were involved in metabolic pathways such as purine metabolism, urea cycle and glutamate metabolism. The link between NADSYN1 and nicotinamide metabolism points to a more extensive impact on cellular redox equilibrium. [8,10] The function of vitamin D-related genes in maintaining lipid homeostasis is further supported by the enrichment of lipid metabolism pathways such as glycerolipid metabolism and cardiolipin biosynthesis.[13]

This enrichment method finds new regulatory elements like transcription factors and miRNAs, that could affect the expression of genes linked to vitamin D deficiency. By connecting genetic results to biological pathways and cellular functions, enrichment analysis offers mechanistic insights in contrast to normal GWAS, which finds variants associated with illness risk. By including metabolomic data, genetic differences and disturbances in vitamin D metabolism are further linked. By relating genetic changes to metabolic processes, this work provides a thorough system biology viewpoint.

**Table 13**
**RNAseq_Automatic_GEO_Signatures_Human_Up**

| Term | Overlap | P-value | Adjusted P-value | Old P-value | Old Adjusted P-value | Odds Ratio | Combined Score | Genes |
|---|---|---|---|---|---|---|---|---|
| Modeling Gata6 Haplo insufficiency Development GSE92581 1 | 2/250 | 9.18 E-04 | 0. 045614 | 0 | 0 | 79. 62903 | 556. 835542 | CALCB; GC |
| Following Depletion Loxl2 Esophgeal GSE135510 1 | 2/250 | 9.18 E-04 | 0. 045614 | 0 | 0 | 79. 62903 | 556. 835542 | CALCB; CYP2R1 |
| Hepatitis Ns5A Stabilizes Growth-Regulatory GSE102910 1 | 2/250 | 9.18 E-04 | 0. 045614 | 0 | 0 | 79. 62903 | 556. 835542 | CALCB; GC |
| Transcriptional Cell-Derived Polarized Hepatocytes GSE123462 1 | 1/250 | 0. 049074 | 0. 049074 | 0 | 0 | 26. 43507 | 79. 68675066 | GC |
| Neutrophils Generalised Pustular Psoriasis GSE123785 1 | 1/250 | 0. 049074 | 0. 049074 | 0 | 0 | 26. 43507 | 79. 68675066 | NADSYN1 |
| Corneas Keratoconus Groups Nrf2-Antioxidant GSE151631 1 | 1/250 | 0. 049074 | 0. 049074 | 0 | 0 | 26. 43507 | 79. 68675066 | CYP2R1 |
| Estrogen-Receptor Mdm2 Inhibition Cdk4 GSE140758 7 | 1/250 | 0. 049074 | 0. 049074 | 0 | 0 | 26. 43507 | 79. 68675066 | NADSYN1 |
| Suv39H1-Low Migratory Populations Cervical GSE103792 1 | 1/250 | 0. 049074 | 0. 049074 | 0 | 0 | 26. 43507 | 79. 68675066 | NADSYN1 |
| Rna-Binding Acinus Component Junction GSE81460 1 | 1/250 | 0. 049074 | 0. 049074 | 0 | 0 | 26. 43507 | 79. 68675066 | NADSYN1 |
| Dmso Epz6438 Cd34 Hspcs GSE144131 1 | 1/250 | 0. 049074 | 0. 049074 | 0 | 0 | 26. 43507 | 79. 68675066 | NADSYN1 |

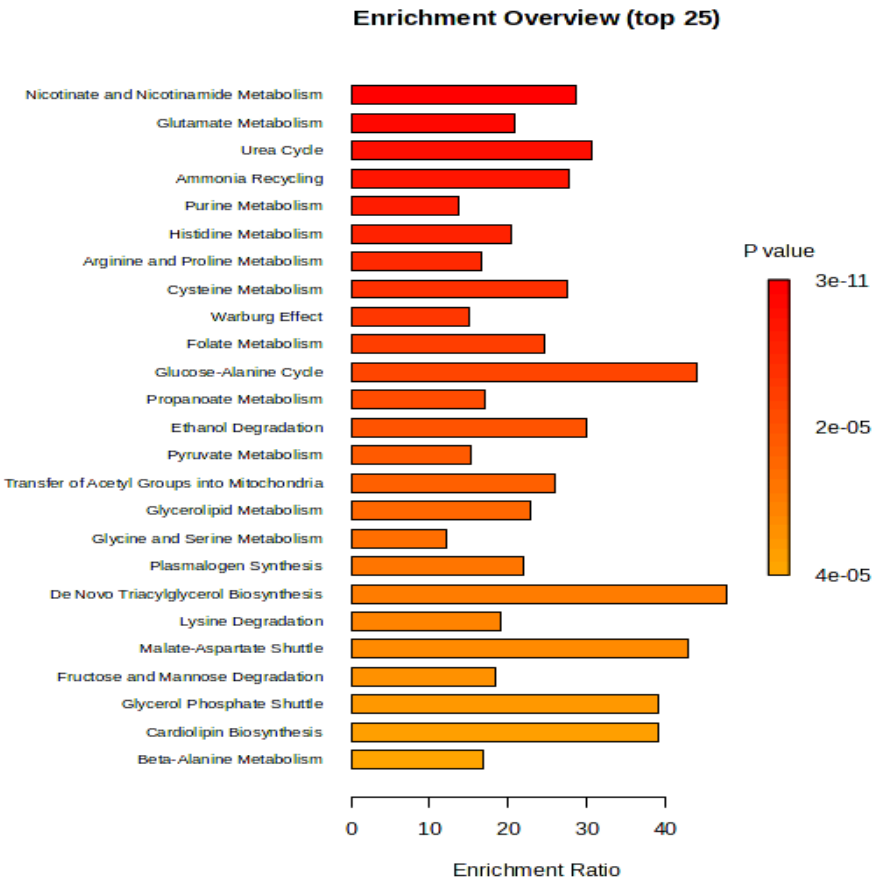


**Figure 1: The top 25 enriched metabolic pathways are shown according to their enrichment ratio. A color gradient denoting statistical significance (p-value) is shown, with darker red denoting more importance.**

This method improves GWAS results and helps precision medicine techniques for identifying and treating vitamin D deficiency by revealing biological pathways and regulatory networks.

## Conclusion

Through a secondary bioinformatics analysis of GWAS data from Wang et al[16], this work seeks to uncover genetic variations linked to vitamin D insufficiency. The results could help with risk assessment, guide individualized preventative plans and pinpoint possible treatment targets for diseases linked to vitamin D insufficiency. At the end, this study will improve knowledge of the genetic makeup of vitamin D metabolism, which will lead to better public health initiatives.

## Acknowledgement

Authors thank Central Research Laboratory for Molecular Genetics, Bioinformatics and Machine Learning at Apollo Institute of Medical Sciences and Research Chittoor Murukamabttu - 571727, Andhra Pradesh, India for the infrastructure.

## References

1. Alcalá-Santiago Á. et al, Vitamin D deficiency and COVID-19: A biological database study on pathways and gene-disease associations, *Int J Mol Sci*, **23**, 14256 **(2022)**

2. Asghari G., Yuzbashian E., Najd-Hassan-Bonab L., Mirmiran P., Daneshpour M.S. and Azizi F., Association of rs2282679 polymorphism in vitamin D binding protein gene (GC) with the risk of vitamin D deficiency in an Iranian population: season-specific vitamin D status, *BMC Endocr Disord*, **23**, 217 **(2023)**

3. Batai K. et al, Common vitamin D pathway gene variants reveal contrasting effects on serum vitamin D levels in African Americans and European Americans, *Hum Genet*, **133**, 1395–1405 **(2014)**

4. Ding Y. et al, Is vitamin D deficiency the cause or the effect of systemic lupus erythematosus: Evidence from bidirectional Mendelian randomization analysis, *J Immunol Res*, **2022**, 8689777 **(2022)**

5. Engelman C.D. et al, Vitamin D intake and season modify the effects of the GC and CYP2R1 genes on 25-hydroxyvitamin D concentrations, *J Nutr*, **143**, 17–26 **(2013)**

6. Hendi N.N., Al-Sarraj Y., Ismail Umlai U.K., Suhre K., Nemer G. and Albagha O., Genetic determinants of vitamin D deficiency in the Middle Eastern Qatari population: a genome-wide association study, *Front Nutr*, **10**, 1242257 **(2023)**

7. Lasky-Su J. et al, Genome-wide association analysis of circulating vitamin D levels in children with asthma, *Hum Genet*, **131**, 1495–1505 **(2012)**

8. Lu Z., Jiao Y. and Li J., Higher genetically predicted triglycerides, LDL and HDL increase the vitamin D deficiency: A Mendelian randomization study, *Front Nutr*, **9**, 862942 **(2022)**

9. Manousaki D. et al, Vitamin D levels and risk of type 1 diabetes: A Mendelian randomization study, *PLoS Med*, **18**, e1003536 **(2021)**

10. Meena D. et al, Associations of genetically predicted vitamin D status and deficiency with the risk of carotid artery plaque: a Mendelian randomization study, *Sci Rep*, **14**, 14743 **(2024)**

11. Moy K.A. et al, Genome-wide association study of circulating vitamin D–binding protein, *Am J Clin Nutr*, **99**, 1424–1431 **(2014)**

12. O'Brien K.M., Sandler D.P., Shi M., Harmon Q.E., Taylor J.A. and Weinberg C.R., Genome-wide association study of serum 25-hydroxyvitamin D in US women, *Front Genet*, **9**, 67 **(2018)**

13. Rivera-Paredez B. et al, Association between vitamin D deficiency and single nucleotide polymorphisms in the vitamin D receptor and GC genes and analysis of their distribution in Mexican postmenopausal women, *Nutrients*, **10**, 1175 **(2018)**

14. Saurabh K., Bala J. and Hemang, Vitamin D status among overweight and obese adolescents, *Biomed Biotechnol Res J*, **4**, 162-165 **(2020)**

15. Silverio C. de S. and Bonilla C., Vitamin D-associated genetic variants in the Brazilian population: Investigating potential instruments for Mendelian randomization, *Biomédica*, **44**, 45–53 **(2024)**

16. Wang T.J. et al, Common genetic determinants of vitamin D insufficiency: a genome-wide association study, *Lancet*, **376**, 180–188 **(2010)**.

(Received 20th March 2025, accepted 15th May 2025)