

IDENTIFICATION AND VISUALIZATION OF CRUCIAL GENES INVOLVED IN HYPOTHYROIDISM AND ANALYZING NETWORK USING CYTOSCAPE

Submitted in partial fulfilment of the requirements for the award of

MASTER OF SCIENCE IN BIOINFORMATICS AND DATA SCIENCE

By

GOKUL S (42739010)



**DEPARTMENT OF BIOINFORMATICS
SCHOOL OF BIO AND CHEMICAL ENGINEERING**

SATHYABAMA

**INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)**

Accredited with Grade "A++" by NAAC | 12B Status by UGC | Approved by AICTE

JEPPIAAR NAGAR, RAJIV GANDHI SALAI, CHENNAI - 600 119

APRIL – 2024



SATHYABAMA

INSTITUTE OF SCIENCE AND TECHNOLOGY

(DEEMED TO BE UNIVERSITY)

Accredited with "A++" grade by NAAC | 12B Status by UGC | Approved by AICTE

Jeppiaar Nagar, Rajiv Gandhi Salai, Chennai – 600 119

www.sathyabama.ac.in

DEPARTMENT OF BIOINFORMATICS

BONAFIDE CERTIFICATE

This is to certify that this Project Report is the bonafide work of **GOKUL S (42739010)**, who carried out the final year project entitled "**IDENTIFICATION AND VISUALIZATION OF CRUCIAL GENES INVOLVED IN HYPOTHYROIDISM AND ANALYZING NETWORK USING CYTOSCAPE**" under our supervision from September 2023 to April 2024.

Dr. DANIEL ALEX ANAND
Internal Guide

Dr. H. JEMMY CHRISTY
Head of the Department

Submitted for Viva Voce Examination held on 20/04/2024

Internal Examiner

External Examiner

DECLARATION

I **GOKUL S (42739010)** hereby declare that the Project Report entitled **“IDENTIFICATION AND VISUALIZATION OF CRUCIAL GENES INVOLVED IN HYPOTHYROIDISM AND ANALYZING NETWORK USING CYTOSCAPE”** done by me under the guidance of Dr. DANIEL ALEX ANAND at Sathyabama Institute of Science and Technology and is submitted in partial fulfillment of the requirements for the award of Master of Science in Bioinformatics and Data Science.

DATE: 20/4/2024

PLACE: CHENNAI



SIGNATURE OF THE CANDIDATE

ACKNOWLEDGEMENTS

I am pleased to acknowledge my sincere thanks to the **Board of Management of SATHYABAMA INSTITUTE OF SCIENCE AND TECHNOLOGY** for their kind encouragement in doing this project for completing it successfully. I am grateful to them.

I extend my sincere gratitude to **Dr. REKHA CHAKRAVARTHY, Dean of the School of Science and Humanities**, and I am deeply grateful for her mentorship and leadership, which have been a constant source of inspiration. Her dedication to fostering a conducive academic environment has been instrumental in the successful completion of this thesis.

I convey my thanks to **DR. H. JEMMY CHRISTY, Head of the Department, Department of Bioinformatics**, for providing me necessary support and details at the right time during the progressive reviews.

I would like to express my sincere and deep sense of gratitude to my Project Guide **Dr. DANIEL ALEX ANAND** for his valuable guidance, suggestions and constant encouragement paved way for the successful completion of my project work.

I wish to express my thanks to all Teaching and Non-teaching staff members of the Department of Bioinformatics who were helpful in many ways for the completion of the project.

ABSTRACT

Hypothyroidism, a disorder characterized by insufficient thyroid hormone production, can cause significant health issues. This study aimed to elucidate the complex genetic networks underlying hypothyroidism. We integrated multiple bioinformatics approaches to identify crucial genes and pathways involved in this condition. Initial dataset compilation from the Alliance of Genome Resources, DisGeNET, and NCBI-GEO yielded 1018 genes associated with hypothyroidism. After filtering, 152 unique genes were analyzed using the DAVID tool for functional enrichment analysis. Subsequently, the STRING database was employed for protein-protein interaction network construction, revealing 129 nodes and 960 edges. Network visualization and analysis in Cytoscape, using the MCODE plugin, identified two significant subnetworks. Cluster two demonstrated disruption of thyroid hormone production and signaling pathways, warranting further investigation. The ClueGO plugin was used for functional enrichment analysis within this subnetwork. Finally, the cytoHubba plugin pinpointed ten hub genes potentially central to hypothyroidism pathogenesis: TG, IFNG, TSHR, SLC5A5, SLC16A2, TSHB, THRB, DIO2, IYD, and THRA. These findings offer valuable insights into the molecular mechanisms of hypothyroidism. The identified hub genes could serve as promising targets for diagnostic and therapeutic interventions. Further research is needed to validate these findings and explore their clinical implications.

Keywords: Hypothyroidism, gene identification, bioinformatics, network analysis, Cytoscape, DAVID, STRING, MCODE, ClueGO, cytoHubba, hub genes.

TABLE OF CONTENT

CHAPTER NO	TITLE	PAGE NO
	ABSTRACT	i
	LIST OF TABLES	v
	LIST OF FIGURES	vi
1	INTRODUCTION	1
1.1	BACKGROUND	2
1.2	EXISTING PROBLEMS	1
1.3	TREATMENT	3
1.4	BOTTLENECK AREAS	4
2	REVIEW OF LITERATURE	5
3	AIM AND SCOPE	8
3.1	AIM	8
3.2	SCOPE	8
4	METHODOLOGY	9
4.1	GENE DATASETS	9
4.2	DATABASES	9
4.2.1	ALLIANCE OF GENOME RESOURCE	9
4.2.2	DISGENET	10
4.2.3	NCBI-GEO	10
4.3	FILTERING THE GENES	11
4.4	FUNCTIONAL ENRICHMENT ANALYSIS (DAVID TOOL)	12
4.5	GENE NETWORK ANALYSIS	12
4.5.1	STRINGdb	12
4.6	CYTOSCAPE - NETWORK ANALYST	13
4.7	SUB-NETWORK ANALYSIS USING MCODE	14

4.8	FUNCTIONAL AND PATHWAY ANALYSIS USING CLUEGO	14
4.9	IDENTIFICATION OF HUBGENES USING CYTOSCAPE	14
5	RESULT AND DISCUSSION	15
5.1	FUNCTIONAL ENRICHMENT ANALYSIS	15
5.1.1	BIOLOGICAL PROCESS	15
5.1.2	MOLECULAR FUNCTION	16
5.1.3	KEGG PATHWAY	17
5.1.4	CELLULAR COMPONENT	19
5.2	STRING - PROTEIN NETWORK	20
5.2.1	COUNT OF GENES INVOLVED IN HYPOTHYROIDISM	21
5.2.2	GENES INVOLVED IN HYPOTHYROIDISM	22
5.2.3	NETWORK SHOWING THE HIGHLIGHTED GENES INVOLVED IN THE PATHWAY OF HYPOTHYROIDISM	22
5.3	CYTOSCAPE	23
5.3.1	MCODE- IDENTIFICATION OF SUBNETWORK	24
5.3.2	CLUEGO – FUNCTIONAL ENRICHMENT ANALYSIS OF SUBNETWORK	28
5.3.2.1	BIOLOGICAL PROCESS OF CLUSTER 1	28
5.3.2.2	KEGG PATHWAY OF CLUSTER 1	29
5.3.2.3	BIOLOGICAL PROCESS OF CLUSTER 2	31
5.3.2.4	KEGG PATHWAY OF CLUSTER 2	32

5.3.3	CYTOHUBBA- IDENTIFICATION OF HUB GENES	33
6	SUMMARY AND CONCLUSION	39
6.1	SUMMARY AND CONCLUSION	39
6.2	CONCLUSION	40
	REFERENCES	41

LIST OF TABLES

TABLE NO	DESCRIPTION	PAGE NO
1	TOP GENES/PROTEINS BASED ON DEGREE	35
2	MCODE SUBNETWORK – CLUSTER 1	36
3	MCODE SUBNETWORK – CLUSTER 2	37
4	HUBGENES	38

LIST OF FIGURES

FIGURE NO	DESCRIPTION	PAGE NO
5.1	GENES ASSOCIATE WITH BIOLOGICAL PROCESS	15
5.2	COMBINED PLOT OF BIOLOGICAL PROCESS	15
5.3	GENES ASSOCIATE WITH MOLECULAR FUNCTION	16
5.4	COMBINED PLOT OF MOLECULAR FUNCTION	16
5.5	GENES ASSOCIATE WITH KEGG PATHWAY	17
5.6	COMBINED PLOT OF KEGG PATHWAY	18
5.7	GENES ASSOCIATE WITH CELLULAR COMPONENTS	19
5.8	COMBINED PLOT OF CELLULAR COMPONENTS	19
5.9	PPI – NETWORK USING STRING DATABASE	21
5.10	DISEASE - GENE ASSOCIATION	21
5.11	NETWORK SHOWING THE HIGHLIGHTED GENES INVOLVED IN HYPOTHYROIDISM	22
5.12	PPI - NETWORK USING CYTOSCAPE	22
5.13	STATS USING ANALYZER	23
5.14	TOP GENES IDENTIFIED USING ANALYZER	23
5.15	MCODE SUBNETWORK - CLUSTER 1	24

5.16	GENES/PROTEINS ASSOCIATED IN CLUSTER 1	25
5.17	MCODE SUBNETWORK - CLUSTER 2	26
5.18	GENES/PROTEINS ASSOCIATED WITH CLUSTER2	27
5.19	RESULT OF CLUEGO PLUGIN NETWORK HIGHLIGHTS	28
5.20	BIOLOGICAL PROCESS IN CLUSTER 1	28
5.21	RESULT OF KEGG PATHWAY USING CLUEGO	29
5.22	NETWORK HIGHLIGHTS KEGG PATHWAYS IN CLUSTER 1	29
5.23	RESULT OF KEGG PATHWAY ANALYSIS USING CLUEGO	30
5.24	RESULT OF CLUEGO PLUGIN FOR CLUSTER 2 NETWORK HIGHLIGHTS	30
5.25	BIOLOGICAL PROCESS IN CLUSTER 2	31
5.26	RESULT OF BIOLOGICAL PROCESS FOR CLUSTER 2	31
5.27	RESULT OF KEGG PATHWAY USING CLUEGO	32
5.28	RESULT OF KEGG PATHWAY ANALYSIS FOR CLUSTER 2	32
5.29	HUBGENES AND SCORE	33

CHAPTER 1

INTRODUCTION

Hypothyroidism, a condition characterized by inadequate thyroid hormone production, represents a significant global health concern. Its prevalence increases with age and is substantially higher in women than in men(Chiovato et al., 2019). The clinical manifestations of hypothyroidism encompass a wide spectrum, ranging from subtle symptoms to overt metabolic derangements. Common presentations include fatigue, weight gain, cold intolerance, depression, cognitive impairment, and an increased risk of cardiovascular disease(Garber et al., 2012).

The etiology of hypothyroidism is multifaceted, with both genetic and environmental factors playing crucial roles. Hashimoto's thyroiditis, an autoimmune condition where the body attacks the thyroid gland, is a leading cause of hypothyroidism(Chaker et al., 2022). Other contributors include iodine deficiency, medical treatments such as radioactive iodine therapy or certain medications, and genetic predisposition (Wiersinga, 2014).

Current standard treatment for hypothyroidism involves lifelong thyroid hormone replacement therapy with levothyroxine (Lu et al., 2024). While effective for the majority of patients, a subset of individuals may experience persistent symptoms or difficulties in achieving optimal thyroid hormone levels despite treatment. Furthermore, current therapies primarily address the resultant hormonal deficiency rather than targeting the underlying causes of thyroid dysfunction(Chiovato et al., 2019).

A growing body of research aims to elucidate the intricate genetic architecture of hypothyroidism. Genome-wide association studies (GWAS) have identified several susceptibility loci, suggesting a polygenic contribution to disease development. However, a comprehensive understanding of the complex network of genes and pathways involved in hypothyroidism remains elusive(Chiovato et al., 2019; Garber et al., 2012).

Delineating the key molecular players in hypothyroidism pathogenesis holds immense potential for transforming various aspects of disease management. Identifying individuals at high risk based on genetic profiles could facilitate early diagnosis and potential preventive interventions (Dai & Sternberg, 2004). Additionally, the discovery of novel genetic targets could revolutionize treatment paradigms, leading to the development of targeted therapies that modulate the underlying disease mechanisms rather than solely providing hormone replacement (Effraimidis & Wiersinga, 2014).

1.1 BACKGROUND

A Master Regulator The thyroid gland, a butterfly-shaped organ located at the base of the neck, holds immense responsibility for regulating numerous vital bodily processes. Its primary function involves the production of thyroid hormones, triiodothyronine (T3) and thyroxine (T4), which act as master switches, influencing metabolism, energy expenditure, growth, development, and the function of virtually every organ system (Chiovato et al., 2019).

A State of Hormonal Imbalance Hypothyroidism arises when the thyroid gland fails to produce sufficient levels of thyroid hormones. This disrupts the delicate hormonal balance, leading to a widespread slowdown of physiological functions and manifesting a diverse range of clinical symptoms.

The prevalence of hypothyroidism is substantial, particularly increasing with age and showing a greater propensity to affect women.

(Chiovato et al., 2019; Qiu et al., 2021) **From Subtle to Severe** The clinical presentation of hypothyroidism is often insidious, with symptoms that may overlap with other conditions, posing a diagnostic challenge. Common manifestations include fatigue, weight gain, cold sensitivity, constipation, dry skin, hair loss, muscle weakness, depression, impaired memory, and in severe cases, slowed heart rate and mental processes.

Left untreated, hypothyroidism can significantly increase the risk of cardiovascular disease, infertility, and cognitive decline.

1.2 EXISTING PROBLEMS

Diagnosis: While blood tests can detect hypothyroidism, identifying the underlying cause (such as autoimmune disease, iodine deficiency, or medication side effects) can be more complex. This can delay appropriate treatment.

Subclinical Hypothyroidism: This milder form of hypothyroidism often has subtle symptoms, making diagnosis and decisions about treatment challenging.

Treatment Variability: Individuals respond differently to thyroid hormone replacement therapy, requiring careful dosage adjustments and monitoring.

Persistent Symptoms: Despite treatment, some individuals continue to experience symptoms like fatigue, potentially indicating a need for combination therapies or further investigating underlying causes.

Unanswered Questions: Addressing the Root Cause The current treatment paradigm primarily focuses on addressing the hormonal deficiency in hypothyroidism. However, understanding the root causes triggering thyroid dysfunction remains an area of active investigation. Both genetic predisposition and environmental triggers, such as iodine deficiency, autoimmune processes, or exposure to certain toxins, have been implicated in the development of hypothyroidism.

1.3 TREATMENT

Levothyroxine: The standard treatment is levothyroxine, a synthetic thyroid hormone replacement. Dosage is individualized based on regular blood tests.

Combination Therapies: In some cases, combining levothyroxine with liothyronine (T3) may be considered, but this approach remains somewhat controversial.

Addressing Underlying Causes: Treatment may also involve addressing iodine deficiency or managing autoimmune conditions like Hashimoto's thyroiditis.

1.4 BOTTLENECK AREAS

Bioinformatics Tools: Depending on the complexity of the analysis, the computational power needed to process and visualize large networks could become a bottleneck. Using efficient tools and potentially cloud-based resources can mitigate this.

Hub Gene Validation: Once hub genes are identified, experimental validation (e.g., through gene expression studies or functional assays) is essential. This step can be time-consuming and resource-intensive.

Translating Network Insights to Clinical Practice: While identifying key genes and networks is valuable, developing diagnostic tools or targeted therapies based on this information involves a long and complex process of validation, safety testing, and clinical trials.

Beyond Hormone Replacement: the Search for Targeted Therapies The current reliance on thyroid hormone replacement underscores the need for therapies that directly target the underlying molecular pathophysiology of hypothyroidism. Identifying key genes and pathways involved in thyroid dysfunction holds the potential to unlock novel therapeutic targets.

Personalized Medicine: The Future of Hypothyroidism Treatment Response to levothyroxine treatment can vary between individuals. Some factors contributing to this variability include differences in drug absorption, metabolism, and underlying genetic polymorphisms. Understanding the genetic factors influencing drug response could pave the way for personalized medicine in hypothyroidism. Pharmacogenomic approaches could enable tailoring of treatment regimens and dosage based on a patient's unique genetic profile, optimizing therapeutic outcomes.

CHAPTER 2

REVIEW OF LITERATURE

The primary purpose of a literature review is to provide a comprehensive understanding of the existing knowledge on a particular topic or subject. It helps to identify gaps in the current knowledge, highlight areas for further research, and establish a theoretical framework for the study.

A literature review typically involves a thorough search of relevant literature and research databases, followed by a careful analysis of the findings. The review may be conducted as part of a larger research project, or as a standalone piece of work, such as a dissertation or thesis.

The following are the various references reviewed for the purpose of the current study.

In the research paper of Hypothyroidism in Context: Where We've Been and Where We're Going by Flavia M, Allan C and Luca C mentioned, Hypothyroidism is a prevalent condition that affects a significant portion of the population worldwide. While iodine deficiency remains a major cause, Hashimoto's disease (chronic autoimmune thyroiditis) is the leading cause in iodine-sufficient regions. Diagnosis relies on biochemical testing, with overt primary hypothyroidism defined by elevated TSH and decreased T4 levels. Symptoms are often non-specific and can negatively impact quality of life.

Levothyroxine, a synthetic T4 replacement, is the primary treatment for hypothyroidism. Despite its widespread use, some individuals continue to experience symptoms, highlighting the need for treatment optimization (Hennessey, 2023) . Future research should focus on accurate prevalence assessment, individualizing levothyroxine dosing, exploring treatment durations, and carefully investigating the potential benefits of combination therapy with T3 (Kishore et al., 2020; Sue & Leung, 2020). Addressing these challenges will enhance our understanding of hypothyroidism and lead to improved, patient-centered treatment approaches.

In Identification of Hub Genes and Pathways Associated with Oxidative Stress of Cartilage in Osteonecrosis of Femoral Head Using Bioinformatics Analysis

Analysis - Osteonecrosis of the femoral head (ONFH) is a debilitating bone disease where oxidative stress plays a critical role in pathogenesis. Bioinformatics analyses have significantly advanced our understanding of the genes and pathways implicated in oxidative stress-induced damage in ONFH (Shi et al., 2022).

Key findings include the identification of hub genes (such as JUN, FOXO3, CASP3, JAK2, RELA, EZH2, ABL1, PTGS2, FBXW7, and MCL1) that exhibit high connectivity and may be central to the disease process. These genes are closely associated with signaling pathways like MAPK, PI3K-AKT-mTOR, and FOXO which regulate cellular stress response, inflammation, and survival.

Transcription factors like TFAP2A, GATA2, SP1, and E2F1 have been implicated in regulating hub gene expression, offering potential targets for modulating oxidative stress in ONFH. Bioinformatics approaches involve accessing gene expression datasets from repositories like GEO, performing differential expression analysis, utilizing tools like DAVID and GSEA for functional enrichment analysis, constructing protein-protein interaction networks with STRING and Cytoscape, and predicting transcription factors through NetworkAnalyst (Shi et al., 2022). These analyses offer promising avenues for biomarker development and therapeutic targeting, requiring further experimental validation and exploration of potential drugs to mitigate oxidative damage and improve ONFH management.

Integrative Analyses of Genes Associated with Hashimoto's Thyroiditis -

Hashimoto's thyroiditis, a common autoimmune disorder, has complex origins that remain to be fully understood. To investigate its molecular mechanisms, this study combined and analyzed two datasets from the GEO database (Qiu et al., 2021; Shi et al., 2022). Differentially expressed genes (DEGs) between Hashimoto's patients and healthy individuals were identified, followed by functional enrichment analysis.

Key findings include 182 DEGs, predominantly upregulated, enriched in pathways related to sensory perception and olfactory function. Protein-protein interaction (PPI) analysis revealed hub genes like KRTAP family members, HTR5A, KNG1, DRD3, HTR1D, TAS2R16, INSL5, TAS2R42, and GRM7. Further investigation

highlighted OTUD4, LLPH, and ECHDC1 as important hub genes potentially regulated by miRNAs. These bioinformatics findings offer insights into Hashimoto's thyroiditis pathogenesis and could pave the way for the exploration of new therapeutic targets, ultimately leading to improved disease management.

Integrative analysis of key candidate genes and signaling pathways in autoimmune thyroid dysfunction related to anti-CTLA-4 therapy by bioinformatics - CTLA-4 inhibitors, while effective in treating cancer, can trigger autoimmune thyroid dysfunction (IRAEs). This study sought to identify key genes and pathways involved in this adverse event. Integrated differentially expressed genes (DEGs) were pinpointed using GEO datasets and text mining. Functional analysis revealed relevant GO terms and KEGG pathways in both hypothyroidism and hyperthyroidism groups. PPI network analysis identified top hub genes: ALB, MAPK1, SPP1, PPARG, and MIF for hypothyroidism and ALB, FCGR2B, CD44, LCN2, and CD74 for hyperthyroidism. These findings offer potential biomarkers for diagnosing and managing IRAEs in cancer patients undergoing CTLA-4 inhibitor therapy (Zhang et al., 2020).

Bioinformatics analysis of key genes and pathways in Hashimoto thyroiditis tissues - Hashimoto thyroiditis (HT), a prevalent autoimmune disease with rising incidence, can lead to thyroid dysfunction and increased cancer risks. Current treatments are limited due to the complex immune processes involved in HT (Efthymidis & Wiersinga, 2014). This study analyzed the dataset GSE29315 to address this knowledge gap. A total of 85 differentially expressed genes (DEGs) were found, mainly involved in immune and inflammatory responses, as well as cytokine interaction and cytotoxicity pathways. Ten hub genes were identified, including IFN- γ , IFN- α , and those in the IL6/JAK/STAT3 pathway, suggesting their potential role in HT development. This bioinformatics analysis highlights the value of exploring DEGs and pathways for understanding HT pathogenesis and identifying potential targets for prevention and treatment (Zheng et al., 2020) .

CHAPTER 3

AIM AND SCOPE

3.1 AIM

This project aims to comprehensively identify and visualize crucial genes involved in hypothyroidism, exploring their functional relationships and potential as therapeutic targets.

Further, project aims to systematically identify the key genes and pathways that drive hypothyroidism. By integrating data from multiple sources and employing powerful bioinformatics tools, this study seeks to illuminate the underlying mechanisms of the disease and pave the way for improved diagnostics and treatments.

3.2 SCOPE

The scope of this project utilizes a multi-database approach to gather comprehensive gene expression data related to hypothyroidism, including public repositories like NCBI GEO, DisGeNET, and the Alliance of Genome Resources. Data mining and filtering strategies will be employed to identify genes consistently associated with hypothyroidism across these databases. Tools like DAVID will analyze these genes to reveal their biological functions, pathways, and potential roles in the disease.

The STRING database will construct a protein-protein interaction (PPI) network to visualize gene connections. This PPI network will be imported into Cytoscape software, where plugins like MCODE will identify key subnetworks potentially involved in hypothyroidism.

CytoHubba will pinpoint hub genes exhibiting high connectivity, highlighting their potential importance. The identified subnetworks and hub genes will be carefully analyzed for their functional significance in hypothyroidism, offering insights into potential novel therapeutic targets or diagnostic biomarker.

CHAPTER 4

METHODOLOGY

4.1 GENE DATA SETS

Gene datasets were downloaded from multiple databases, and the vast data was filtered according to our parameters using Microsoft Excel.

4.2 DATABASES

4.2.1 ALLIANCE OF GENOME RESOURCE - www.alliancegenome.org

The Alliance of Genome Resources is a consortium that brings together major model organism databases and the Gene Ontology resource. Their goal is to create a central platform for researchers to explore genes across humans and well-studied model organisms (Kishore et al., 2020). This platform will integrate data from various sources using common formats, making it easier to compare and analyze information.

Key points:

- Launched in 2016, the Alliance provides a central portal for accessing data on model organisms, gene ontology, and human data.
- All data types are open and freely available.

The Alliance aims to expand coverage to include more model organisms and data types relevant to non-model-organism research, ultimately supporting discoveries in human health and disease.

For my project, I utilized the Alliance of Genome Resources database to identify genes implicated in hypothyroidism. This powerful resource integrates data from various model organisms and Gene Ontology, providing a comprehensive platform for genetic research. By leveraging the Alliance's curated datasets and tools, I was able to efficiently download a list of relevant genes, which will contribute significantly to my ongoing analysis of the genetic basis of hypothyroidism.

4.2.2 DISGENET – <https://www.disgenet.org/>

DisGeNET emerges as a valuable tool for researchers investigating the genetic basis of human diseases. It addresses critical challenges in the field, such as fragmented data, inconsistent formats, and limited accessibility.

DisGeNET is a powerful platform for exploring genes associated with human diseases. It stands out due to its comprehensive data collection, seamless integration of data from multiple sources, and use of standardized vocabularies for consistency (Piñero et al., 2021). DisGeNET's prioritization tools help researchers identify the most relevant genotype-phenotype associations. The platform's user-friendly web interface, along with access through Cytoscape, SPARQL, and programming tools, ensures accessibility for various research needs. Researchers can utilize DisGeNET to investigate disease mechanisms, analyze disease genes, inform drug discovery processes, and computationally validate findings.

To investigate the genetic underpinnings of hypothyroidism, I employed the DisGeNET database (Piñero et al., 2017). This comprehensive resource integrates curated data on gene-disease associations from various sources, including expert repositories, GWAS catalogs, and scientific literature. Using DisGeNET's search and filtering tools, I was able to identify and download a list of genes potentially involved in hypothyroidism. This dataset serves as a crucial foundation for further analysis within my project.

4.2.3 NCBI – GEO – <http://www.ncbi.nlm.nih.gov/geo/>

The GEO database stands as a pivotal public repository for researchers exploring high-throughput gene expression data and other functional genomics datasets. Since its inception in 2000, GEO has continuously evolved alongside rapidly advancing technologies. It now encompasses diverse data types extending beyond gene expression, such as those relating to genome methylation, chromatin structure, and genome-protein interactions.

GEO's value lies in its adherence to community-derived reporting standards. This ensures the inclusion of essential components like raw data, processed data, and descriptive metadata, guaranteeing the reproducibility and reliability of studies.

Beyond simply storing vast amounts of data, GEO empowers researchers with a suite of user-friendly web-based tools. These tools facilitate targeted searches for datasets relevant to specific interests, along with powerful visualization and analysis capabilities.

I utilized the Gene Expression Omnibus (GEO) database to analyze the dataset GSE176153. This dataset contains gene expression profiles from four hypothyroidism samples and four normal thyroid samples. My analysis of this GEO dataset will provide valuable insights into the molecular differences between hypothyroidism and healthy thyroid function.

4.3 FILTERING THE GENES

To identify robust candidate genes implicated in hypothyroidism, I conducted a rigorous analysis across multiple gene-disease association databases. I began by querying the Alliance of Genome Resources, NCBI GEO, and DisGeNET, yielding a comprehensive initial list of 1018 potential genes. To increase confidence in the results, I applied a cross-database filtering strategy using Excel. This approach focused on finding genes present in all three databases. The analysis successfully narrowed the list down to 152 high-confidence candidate genes strongly associated with hypothyroidism. These validated genes offer a focused starting point for further investigation into the molecular mechanisms of hypothyroidism and potential therapeutic targets.

Total genes:

Alliance of genome resource : 154

NCBI – GEO : 250

DisGenet - 614

4.4 FUNCTIONAL ENRICHMENT ANALYSIS (DAVID TOOL)

DAVID, a widely used tool for gene list analysis, received major updates in 2021 (Sherman et al., 2022). A key enhancement was the expansion of its knowledge base to cover more organisms (increasing taxonomy coverage from 17,399 to 55,464) (Lu et al., 2024). All existing gene annotations were updated in line with this expanded coverage, significantly augmenting the resources within the knowledgebase.

Additionally, DAVID incorporated several new types of annotations:

- Small molecule-gene interactions (PubChem)
- Drug-gene interactions (DrugBank)
- Tissue expression data (Human Protein Atlas)
- Disease associations (DisGeNET)
- Pathway information (WikiPathways and PathBank)

Following comprehensive gene filtering, I have imported a refined list of 152 genes into the DAVID tool. These genes will now undergo functional enrichment analysis to uncover their potential roles in biological processes, pathways, and diseases.

4.5 GENE NETWORK ANALYSIS

The network analysis was based only on the filtered genes taken from all three databases. STRINGdb, Cytoscape and NetworkAnalyst tools were considered for Network analysis of the genes. Their Networks and interactions have been visualized to better understand disease development.

4.5.1 STRINGdb: <https://string-db.org/>

Understanding the intricate network of protein-protein interactions (PPIs) is crucial for deciphering cellular function. While a wealth of PPI information exists, it often remains fragmented across various sources. The STRING database offers a solution by systematically collecting and integrating both physical and functional PPIs from diverse sources. These sources include text mining of scientific literature, computational predictions, curated databases, and interaction experiments. STRING critically evaluates and scores interactions, even transferring knowledge

to less-studied organisms (Szklarczyk et al., 2023). Key features of STRING v12.0 include the ability to analyze novel genomes, enhanced co-expression prediction, confidence scoring of experimental interactions, and advanced functional enrichment analysis capabilities. STRING provides a comprehensive and powerful resource for researchers exploring the complex world of protein interactions and their implications for cellular processes.

The 152 genes were analyzed for Protein network interactions using STRINGdb. The interactions and the genes involved in the development of the disease and their interactions were visualized.

The STRING network was downloaded in a .TSV file and was imported into Cytoscape for Network visualization , identification of subnetworks and functional and pathway analysis using MCODE and CLUEGO , identification of the hubs present in the network using Cytohubba.

4.6 CYTOSCAPE - NETWORK ANALYST

The network analysis was based only on the filtered genes. STRINGdb, Cytoscape and NetworkAnalyst tools were considered for Network analysis of the genes (Doncheva et al., 2019). Their Networks and interactions have been visualized to better understand disease development. After creating a protein-protein interaction network in the STRING database, I exported it to Cytoscape for in-depth visualization (Shannon et al., 2003). Using Cytoscape's Network Analyzer, calculate key network metrics including Betweenness Centrality, Closeness Centrality, and node degree (Doncheva et al., 2019, 2023). Additionally, determine the number of undirected edges in the network. These analyses will help identify highly connected genes (hubs) within the protein network, which may play crucial roles in biological processes (Bindea et al., 2009; Majeed & Mukhtar, 2023).

4.7 SUB - NETWORK ANALYSIS USING MCODE

MCODE (Molecular Complex Detection) is a popular Cytoscape plugin designed to identify densely connected clusters within large protein-protein interaction (PPI) networks. These clusters often correspond to functional biological modules or molecular complexes (Franz et al., 2023). MCODE is valuable for bioinformatics research because it emphasizes interconnectedness within biological networks, allows for analysis of complex networks, and offers a user-friendly interface within Cytoscape (Bader & Hogue, 2003).

4.8 FUNCTIONAL AND PATHWAY ANALYSIS USING CLUEGO

ClueGO is a Cytoscape plugin that streamlines the analysis of large gene sets. It integrates Gene Ontology (GO), KEGG pathways, and other ontologies to visualize functional relationships between genes (Bindea et al., 2009). ClueGO identifies overrepresented biological processes and pathways within your data, providing clarity and biological insights. Key features include informative network visualizations, rigorous statistical analysis, customization options, and integration with CluePedia for detailed pathway analysis. Overall, ClueGO transforms complex gene lists into meaningful interpretations, offers visually appealing networks for exploration, and provides a user-friendly experience within Cytoscape.

4.9 IDENTIFICATION OF HUBGENES USING CYTOHUBBA

Networks provide a powerful way to represent biological data like protein interactions, gene regulation, and cellular signaling. By analyzing a node's properties within a network, we can identify crucial elements within these biological systems. cytoHubba, a Cytoscape plugin, was developed to facilitate this process (Chin et al., 2014) . It offers 11 different methods for ranking nodes based on their network features. One novel method, MCC, demonstrates superior performance in predicting essential proteins within yeast PPI networks. cytoHubba's user-friendly interface and comprehensive analysis options make it a valuable tool for exploring biological networks. This type of network analysis can lead to valuable insights about essential regulatory networks and potentially identify new protein drug targets. The plugin's popularity is evident from its significant download statistics.

CHAPTER 5

RESULTS AND DISCUSSION

5.1 FUNCTIONAL ENRICHMENT ANALYSIS

I conducted functional enrichment analysis on my refined list of 152 genes using the DAVID online tool. This analysis focused on illuminating the biological processes, cellular components, molecular functions, and KEGG pathways associated with these genes. These results will provide valuable insights into the potential roles these genes play in hypothyroidism.

5.1.1 BIOLOGICAL PROCESS

Category	Term	Count	%	PValue	Genes	List Total	Pop Hits	Pop Total	Fold Enrichment	Bonferroni	Benjamini	FDR
UP_KW_BIOLOGICAL_PROCESS	KW-0893*Thyroid hormones biosynthesis	6	4.651162791	2.77E-10	DUOX1, TPO, TG, DIO1, DIO2, DUOX2	79	7	11371	123.3743219	1.69E-08	1.69E-08	1.69E-08
UP_KW_BIOLOGICAL_PROCESS	KW-0376*Hydrogen peroxide	4	3.100775194	1.70E-05	DUOX1, TPO, CAT, DUOX2	79	8	11371	71.96835443	0.001034305	5.17E-04	5.17E-04
UP_KW_BIOLOGICAL_PROCESS	KW-0090*Biological rhythms	7	5.426356589	5.61E-04	NGFR, EGRI, NCOR1, PPARG, NR1D1, PPARGC1A, CDK5R1	79	150	11371	6.717046414	0.033623712	0.011397467	0.011397467
UP_KW_BIOLOGICAL_PROCESS	KW-0094*Blood coagulation	3	2.325581395	0.046042343	THBD, F7, ANXA5	79	50	11371	8.636202532	0.943600535	0.561716587	0.561716587
UP_KW_BIOLOGICAL_PROCESS	KW-0356*Hemostasis	3	2.325581395	0.046042343	THBD, F7, ANXA5	79	50	11371	8.636202532	0.943600535	0.561716587	0.561716587
UP_KW_BIOLOGICAL_PROCESS	KW-0735*Steroidogenesis	2	1.550387597	0.072950686	STAR, HSD3B1	79	11	11371	26.1703107	0.9901537	0.741665308	0.741665308

Fig 5.1: Genes associate in Biological Process

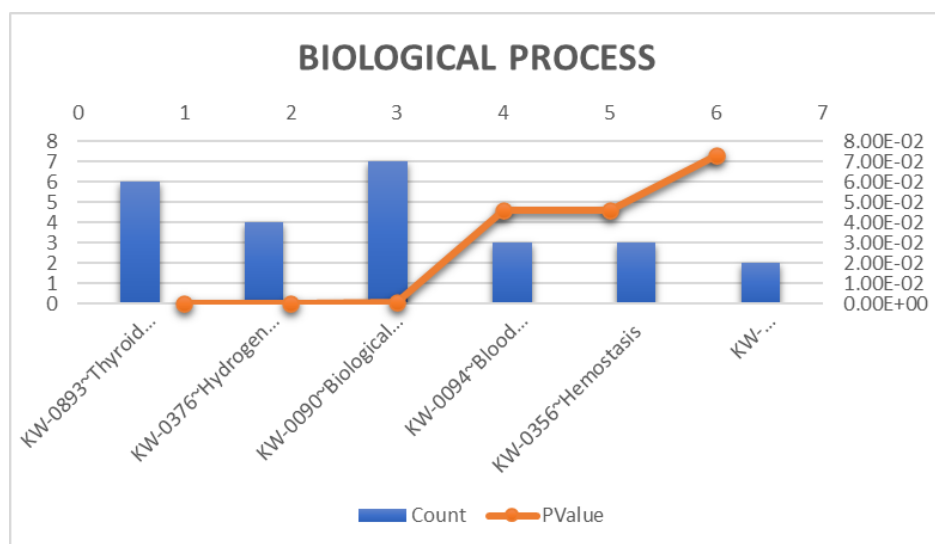


Fig 5.2: Combined Plot of Biological Process

Within the biological processes identified, "Thyroid Hormone Biosynthesis" stands out as the most significant. Six genes are involved in this pathway, demonstrating

a Fold Enrichment of 123.37. This exceptionally high Fold Enrichment, coupled with a very low P-value ($2.77\text{E-}10$), indicates a strong statistical association between these genes and the biological process of Thyroid Hormone Biosynthesis.

5.1.2 MOLECULAR FUNCTION

Category	Term	Count	%	PValue	Genes	List Total	Pop Hits	Pop Total	Fold Enrichment	Bonferroni	Benjamini	FDR
UP_KW_MOLECULAR_FUNCTION	KW-0372~Hormone	13	10.07751898	1.84E-11	CSH2, CSH1, INHBB, TRH, CSHL1, GIP, INS, GH2, POMC, GH1, TG, BGLAP, TSHB	97	98	11834	16.18367947	9.01E-10	9.01E-10	8.82E-10
UP_KW_MOLECULAR_FUNCTION	KW-0560~Oxidoreductase	21	16.27909877	7.36E-08	GAPD, NOS2, NOS3, SROSA1, DIO1, HSD3B1, DIO2, MTHFR, PLOD3, PLOD2, PLOD1, SOD2, CYP19A1, MT-NOS, SOD1, DUOX1, TPO, IYD, CAT, PAM, DUOX2	97	612	11834	4.18627451	3.61E-06	1.80E-06	1.77E-06
UP_KW_MOLECULAR_FUNCTION	KW-0339~Growth factor	7	5.426356589	7.30E-04	IL4, IL6, MSTN, BDNF, INHBB, IGF1, NGF	97	133	11834	6.421052632	0.035128552	0.011915613	0.011672437
UP_KW_MOLECULAR_FUNCTION	KW-0575~Peroxidase	4	3.100775184	0.001045366	DUOX1, TPO, CAT, DUOX2	97	25	11834	19.52	0.049958586	0.012805727	0.012544385
UP_KW_MOLECULAR_FUNCTION	KW-0112~Calmodulin-binding	7	5.426356589	0.001720099	NOS2, NOS3, PHX1A1, MYH6, MYH7, NRGN, SLC9A1	97	157	11834	5.439490446	0.080979113	0.016856917	0.016512951
UP_KW_MOLECULAR_FUNCTION	KW-0202~Cytokine	6	4.651162791	0.021114903	IL10, IL4, IL6, IFNG, MSTN, TNF	97	195	11834	3.753846154	0.648557632	0.172438374	0.168919224
UP_KW_MOLECULAR_FUNCTION	KW-0514~Muscle protein	3	2.325581395	0.095055925	PHX1A1, MYH6, MYH7	97	64	11834	5.71875	0.992510285	0.665391475	0.651812057

Fig 5.3: Genes associate in Molecular Function

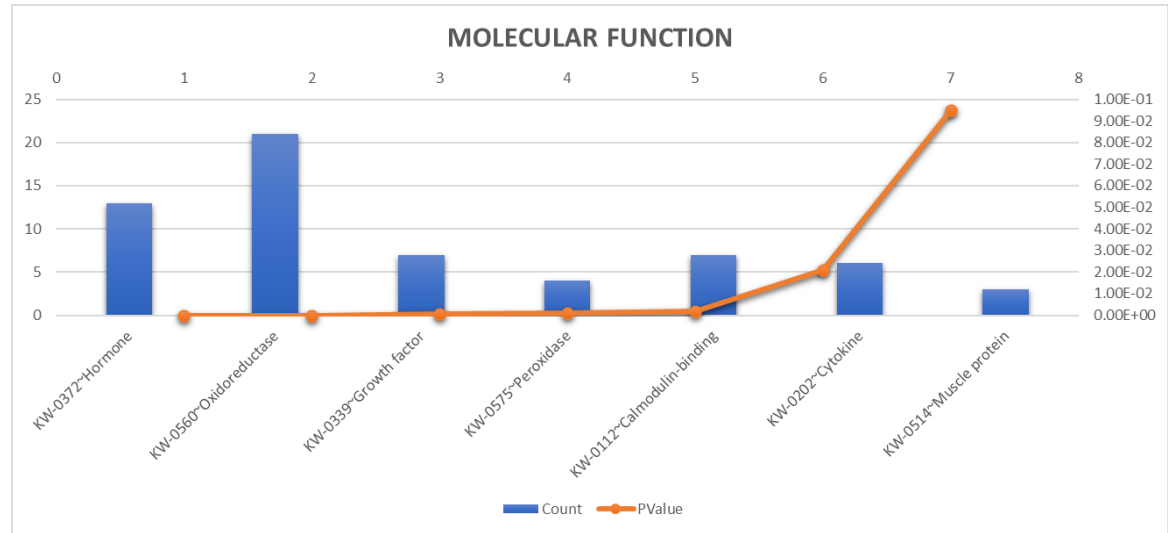


Fig 5.4: Combined Plot of Molecular Function

Oxidoreductase Activity: This function had the highest abundance, with 21 genes involved. It exhibited a Fold Enrichment of 4.19, suggesting these genes are overrepresented compared to what would be expected by chance. The associated P-value of 7.36E-08 indicates statistical significance.

Hormone Activity: The second highest, with 13 associated genes, showed an even stronger enrichment signal (Fold Enrichment of 16.18) and a highly significant P-value (1.84E-11).

Overall, while both functions appear relevant, the results point to a particularly strong association between 'Hormone Activity' and the genes analyzed. This aligns with the study's focus on hypothyroidism, which directly involves hormone dysfunction.

5.1.3 KEGG PATHWAY

Category	Term	Count	%	PValue	Genes	List	Total	Pop	Hits	Pop	Total	Fold	Enrichment	Bonferroni	Benjamini	FDR
KEGG_PATHWAY	hsa04151:PI3K-Akt signaling pathway	17	13.1782946	9.81E-06	NGFR, CSH2, NOS3, BDNF, BAD, CSH1, IGF1, NGF, INS, GH2, GHR, IL4, GH1, IL6, COL2A1, RELN, COL1A2	110	359	8644	3.721144594	0.00219454	5.49E-04	4.66E-04				
KEGG_PATHWAY	hsa05022:Pathways of neurodegeneration - multiple disease	16	12.4031008	8.82E-04	ATP5PO, NOS2, BDNF, WNT3A, BAD, ATP2A2, TNF, MT-NOS, SOD1, IL6, CAT, ATP5PO, NEFL, NEFM, NEFH, CDK5R1	110	476	8644	2.641405653	0.17926103	0.01518947	0.01288392				
KEGG_PATHWAY	hsa04060:Cytokine-cytokine receptor interaction	15	11.627907	1.99E-05	IL10, NGFR, MSTN, CSH2, CSH1, INHBB, NGF, TNF, GH2, GHR, IL4, GH1, IL6, IFNG, LEP	110	297	8644	3.968778696	0.00445667	7.44E-04	6.31E-04				
KEGG_PATHWAY	hsa04080:Neuroactive ligand-receptor interaction	15	11.627907	1.97E-04	THRB, THRA, CSH2, CSH1, TRH, ADRB2, TSHR, TRHR, GIP, GH2, GHR, POMC, GH1, LEP, TSHB	110	367	8644	3.211790934	0.04308693	0.0055048	0.00466925				
KEGG_PATHWAY	hsa04919:Thyroid hormone signaling pathway	13	10.0775194	3.26E-08	NCOA1, THRB, THRA, BAD, DIO1, DIO2, SLC2A1, ATP2A2, SLC9A1, NCOR1, SLC16A2, MYH6, MYH7	110	121	8644	8.442674681	7.29E-06	3.65E-06	3.09E-06				
KEGG_PATHWAY	hsa05010:Alzheimer disease	13	10.0775194	0.00317164	ATP5PO, NOS2, WNT3A, BAD, IRS4, ATP2A2, TNF, MT-NOS, INS, IL6, ATP5PO, APOE, CDK5R1	110	384	8644	2.66032197	0.50913007	0.0349578	0.02965171				
KEGG_PATHWAY	hsa04918:Thyroid hormone synthesis	12	9.30232558	1.92E-09	DUOX1, TPO, TG, PAX8, IYD, ALB, DUOX2, SERPINA7, TSHB, TSHR, DUOX2, SLC5A5	110	75	8644	12.57309091	4.30E-07	4.30E-07	3.65E-07				
KEGG_PATHWAY	hsa05200:Pathways in cancer	12	9.30232558	0.07063514	NCOA1, IL4, IL6, SHH, IFNG, PAX8, NOS2, WNT3A, BAD, SLC2A1, PPARG, IGF1	110	531	8644	1.775860298	0.99999993	0.30505654	0.25875332				
KEGG_PATHWAY	hsa05014:Amyotrophic lateral sclerosis	11	8.52713178	0.0162975	ATP5PO, NOS2, BAD, CAT, ATP5PO, NEFL, NEFM, TNF, NEFH, MT-NOS, SOD1	110	364	8644	2.374725275	0.97479507	0.09606948	0.08148751				
KEGG_PATHWAY	hsa04630:JAK-STAT signaling pathway	10	7.75193798	2.36E-04	GH2, IL10, GHR, IL4, GH1, IL6, IFNG, CSH2, LEP, CSH1	110	166	8644	4.733844469	0.05146134	0.00586961	0.00497869				
KEGG_PATHWAY	hsa05020:Prion disease	10	7.75193798	0.00726041	EGFR1, IL6, ATP5PO, BAD, CAV1, ATP5PO, NCAM1, TNF, MT-NOS, SOD1	110	272	8644	2.889037433	0.80451367	0.05471694	0.04641166				
KEGG_PATHWAY	hsa05415:Diabetic cardiomyopathy	9	6.97674419	0.00403404	G6PD, ATP5PO, COL1A2, NOS3, SLC2A1, ATP5PO, ATP2A2, MT-NOS, INS	110	203	8644	3.483922974	0.59564068	0.03765099	0.03193611				
KEGG_PATHWAY	hsa04024:cAMP signalling pathway	9	6.97674419	0.00742382	POMC, BDNF, BAD, ATP2A2, ADRB2, TSHB, TSHR, SLC9A1, GIP	110	225	8644	3.143272727	0.81159059	0.05471694	0.04641166				
KEGG_PATHWAY	hsa05320:Autoimmune thyroid disease	8	6.20155039	3.99E-06	IL10, IL4, IL6, TPO, TG, HLA-DQB1, TSHB, TSHR, HLA-DQB1	110	53	8644	11.86140652	8.94E-04	2.98E-04	2.53E-04				
KEGG_PATHWAY	hsa05321:inflammatory bowel disease	8	6.20155039	1.59E-05	IL10, IL4, IL6, IFNG, TNF, FOXR3, HLA-DQB1, HLA-DQB1	110	65	8644	9.671608392	0.00356236	7.14E-04	6.05E-04				
KEGG_PATHWAY	hsa04022:cGMP-PKG signaling pathway	8	6.20155039	0.00506975	BAD, NOS3, IRS4, ATP2A2, ADRB2, MYH6, MYH7, INS	110	167	8644	3.764398476	0.67970463	0.04542493	0.03853007				
KEGG_PATHWAY	hsa0152: Tuberculosis	8	6.20155039	0.00757243	IL10, IL6, IFNG, NOS2, BAD, TNF, HLA-DQB1, HLA-DQB1	110	180	8644	3.492525253	0.8178054	0.05471694	0.04641166				
KEGG_PATHWAY	hsa05016:Huntington disease	8	6.20155039	0.0914043	ATP5PO, BDNF, ATP5PO, PPARG, SOD2, MT-NOS, PPARGC1A, SOD1	110	306	8644	2.054426619	1	0.36103698	0.30623671				
KEGG_PATHWAY	hsa05140:Leishmaniasis	7	5.476165659	4.03E-04	IL10, IL4, IFNG, MYD88, TNF, HLA-DQB1, HLA-DQB1	110	77	8644	7.143801655	0.00604473	0.00899956	0.007630716				

Fig 5.5: Genes Associate with KEGG Pathway

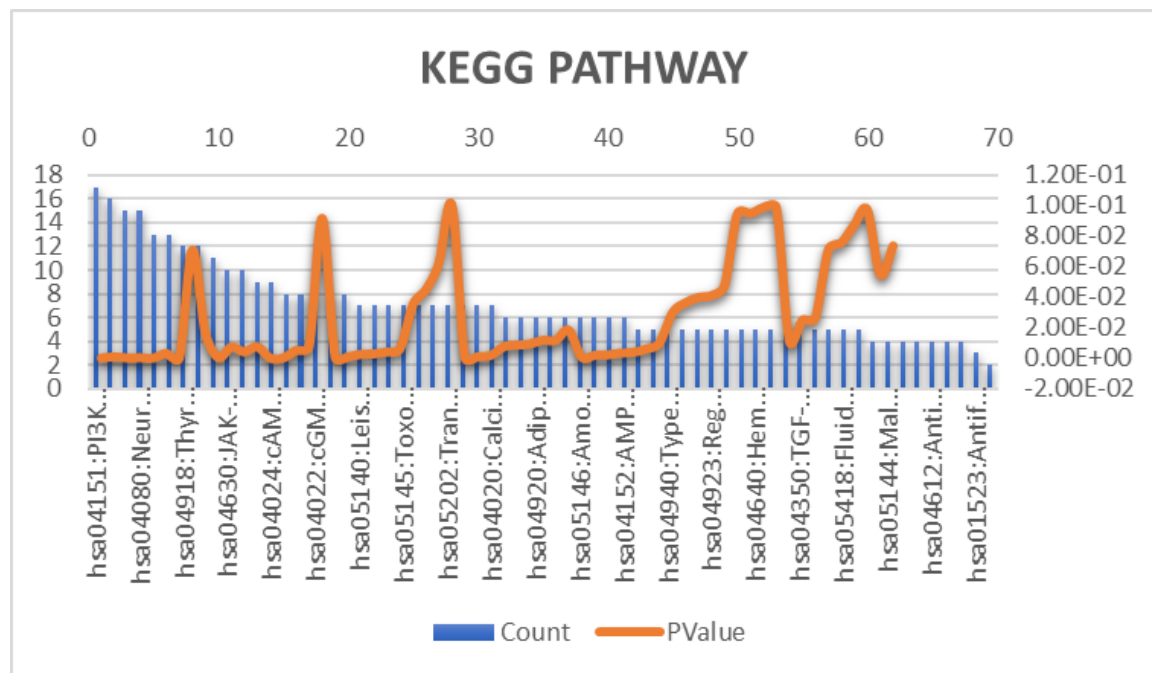


Fig 5.6: Combined plot of KEGG Pathway

My KEGG pathway analysis revealed three pathways significantly associated with my gene set. These pathways, directly related to thyroid function, are:

- Thyroid Hormone Signaling Pathway: 13 genes were implicated, with a Fold Enrichment of 8.44 and a P-value of 3.26E-08.
- Thyroid Hormone Synthesis: This pathway showed the highest Fold Enrichment (12.57) and a very significant P-value (1.92E-09), with 12 associated genes.
- Autoimmune Thyroid Disease: 8 genes were involved with this pathway, demonstrating a Fold Enrichment of 11.86 and a P-value of 3.99E-09.

All three pathways are important, the analysis suggests that disruption of thyroid hormone synthesis may be the most pronounced aspect in this particular case of hypothyroidism. Understanding these differences can help prioritize areas for deeper investigation.

5.1.4 CELLULAR COMPONENT

Category	Term	Count	%	PValue	Genes	List Total	Pop Hits	Pop Total	Fold Enrichment	Bonferroni	Benjamini	FDR
UP_KW_CELLULAR_COMPONENT	KW-0964~Secreted	43	33.33333333	2.79E-10	MSTN, CSH2, CSH1, PLOD3, TRH, TNF, SERPINA	126	2164	17836	2.812795235	9.22E-09	9.22E-09	8.66E-09
UP_KW_CELLULAR_COMPONENT	KW-0162~Chylomicron	3	2.325581395	0.00211345	APOC3, APOE, APOA5	126	10	17836	42.46666667	0.067436149	0.03378559	0.031737978
UP_KW_CELLULAR_COMPONENT	KW-0850~VLDL	3	2.325581395	0.003071417	APOC3, APOE, APOA5	126	12	17836	35.38888889	0.096530333	0.03378559	0.031737978
UP_KW_CELLULAR_COMPONENT	KW-0345~HDL	3	2.325581395	0.00770688	LIPC, APOE, APOA5	126	19	17836	22.35087719	0.225325596	0.06358184	0.059728395
UP_KW_CELLULAR_COMPONENT	KW-0272~Extracellular matrix	6	4.651162791	0.046120854	COL2A1, RELN, COL1A2, ANK2, WNT3A, APOI	126	278	17836	3.055155875	0.789486145	0.304397638	0.285949297

Fig 5.7: Genes Associated with Cellular Component

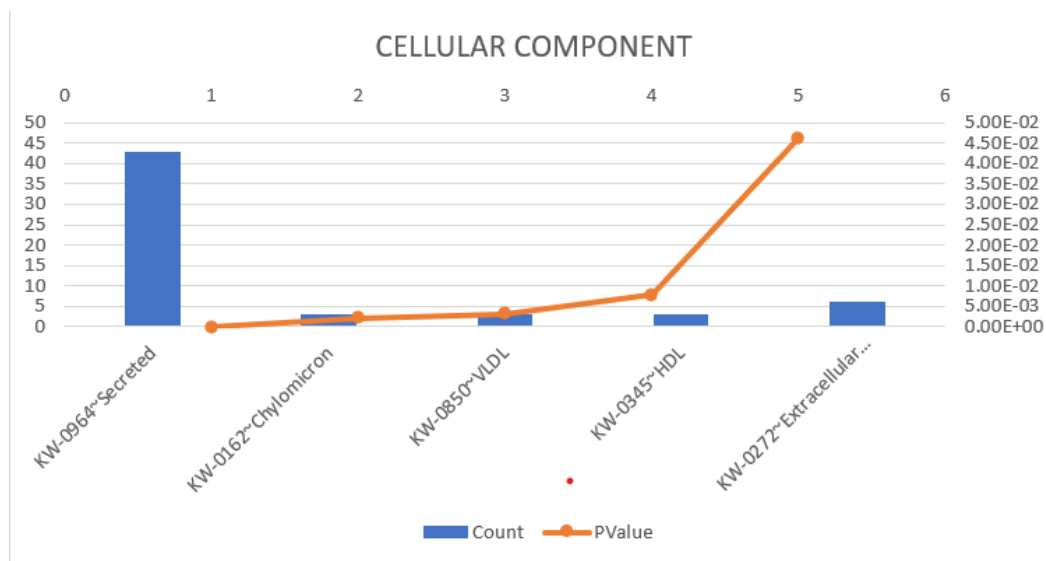


Fig 5.8: Combined plot of Cellular Component

This analysis revealed that the cellular component "Secreted" is overrepresented in my gene set, with 43 associated genes. This finding aligns with the role of the thyroid gland, which secretes hormones essential for regulating metabolism and other bodily functions. Disruptions in the secretion of thyroid hormones are central to hypothyroidism. Understanding the genes and factors regulating the "Secreted" component could reveal bottlenecks in hormone production or release that contribute to the disease.

5.2 STRING - PROTEIN NETWORK

To gain a deeper understanding of the molecular mechanisms underlying hypothyroidism, I used the STRING database to build a network of 152 relevant genes. This analysis focused on pinpointing genes directly involved in the disease and examining how they interact within broader biological pathways.

Network Stats

number of nodes: 133

number of edges: 960

average node degree: 14.4

avg. local clustering coefficient: 0.597

expected number of edges: 333

PPI enrichment p-value: $< 1.0e-16$

PPI enrichment p-value: $< 1.0e-16$ indicates that proteins have more interactions among themselves than expected for a random set of proteins of the same size and degree distribution drawn from the genome. Such enrichment indicates that the proteins are partially biologically connected as a group. It is used to understand the underlying molecular mechanisms of the disease.

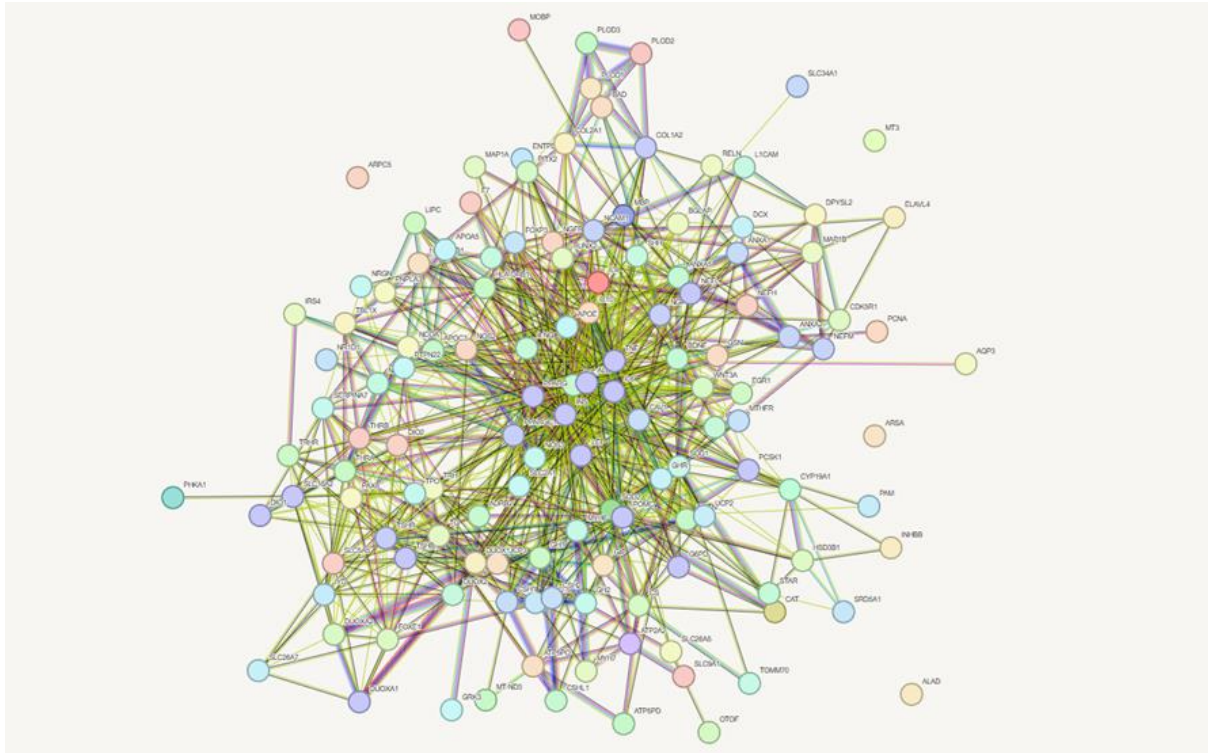


Fig 5.9: PPI – Network using String Database

5.2.1 COUNT OF GENES INVOLVED IN HYPOTHYROIDISM

Disease-gene associations (DISEASES)				
disease	description	count in network	strength	false discovery rate
DOID:7165	Subacute thyroiditis	3 of 3	2.17	0.00032
DOID:3593	Epithelioid trophoblastic tumor	2 of 2	2.17	0.0081
DOID:2856	Euthyroid sick syndrome	2 of 2	2.17	0.0081
DOID:1931	Hypothalamic disease	2 of 2	2.17	0.0081
DOID:14289	Ebstein anomaly	2 of 2	2.17	0.0081
DOID:13208	Background diabetic retinopathy	2 of 2	2.17	0.0081
DOID:13197	Nodular goiter	2 of 2	2.17	0.0081
DOID:12785	Diabetic polyneuropathy	2 of 2	2.17	0.0081
DOID:12175	Dyshormonogenic goiter	2 of 2	2.17	0.0081
DOID:11712	Lipoatrophic diabetes mellitus	2 of 2	2.17	0.0081
DOID:11634	Myxedema	2 of 2	2.17	0.0081
DOID:7997	Thyrotoxicosis	4 of 5	2.07	2.04e-05
DOID:9111	Cutaneous leishmaniasis	3 of 4	2.05	0.00050
DOID:7998	Hyperthyroidism	5 of 7	2.02	1.13e-06
DOID:7188	Autoimmune thyroiditis	6 of 9	1.99	7.41e-08
DOID:12176	Goiter	6 of 9	1.99	7.41e-08
DOID:9521	Laron syndrome	2 of 3	1.99	0.0117
DOID:9370	Exophthalmos	2 of 3	1.99	0.0117
DOID:3596	Placental site trophoblastic tumor	2 of 3	1.99	0.0117
DOID:2942	Bronchiolitis	2 of 3	1.99	0.0117
DOID:2855	Hyperthyroxinemia	2 of 3	1.99	0.0117
DOID:13994	Cleidocranial dysplasia	2 of 3	1.99	0.0117
DOID:13375	Temporal arteritis	2 of 3	1.99	0.0117
DOID:11633	Thyroid hormone resistance syndrome	2 of 3	1.99	0.0117
DOID:11504	Autonomic neuropathy	2 of 3	1.99	0.0117
DOID:0080522	Thyroid gland anaplastic carcinoma	2 of 3	1.99	0.0117
DOID:0050489	Multinodular goiter	2 of 3	1.99	0.0117
DOID:9146	Visceral leishmaniasis	3 of 5	1.95	0.00075
DOID:8283	Peritonitis	3 of 5	1.95	0.00075
DOID:11396	Pulmonary edema	3 of 5	1.95	0.00075
DOID:1459	Hypothyroidism	16 of 28	1.93	3.59e-21
DOID:0050328	Congenital hypothyroidism	15 of 26	1.93	5.57e-20
DOID:6132	Bronchitis	2 of 4	1.87	0.0158
DOID:3491	Turner syndrome	2 of 4	1.87	0.0158
DOID:3113	Papillary carcinoma	2 of 4	1.87	0.0158
DOID:2957	Pulmonary tuberculosis	2 of 4	1.87	0.0158
DOID:14018	Alcoholic liver cirrhosis	2 of 4	1.87	0.0158

Fig 5.10: Disease - gene association

5.2.2 GENES INVOLVED IN HYPOTHYROIDISM

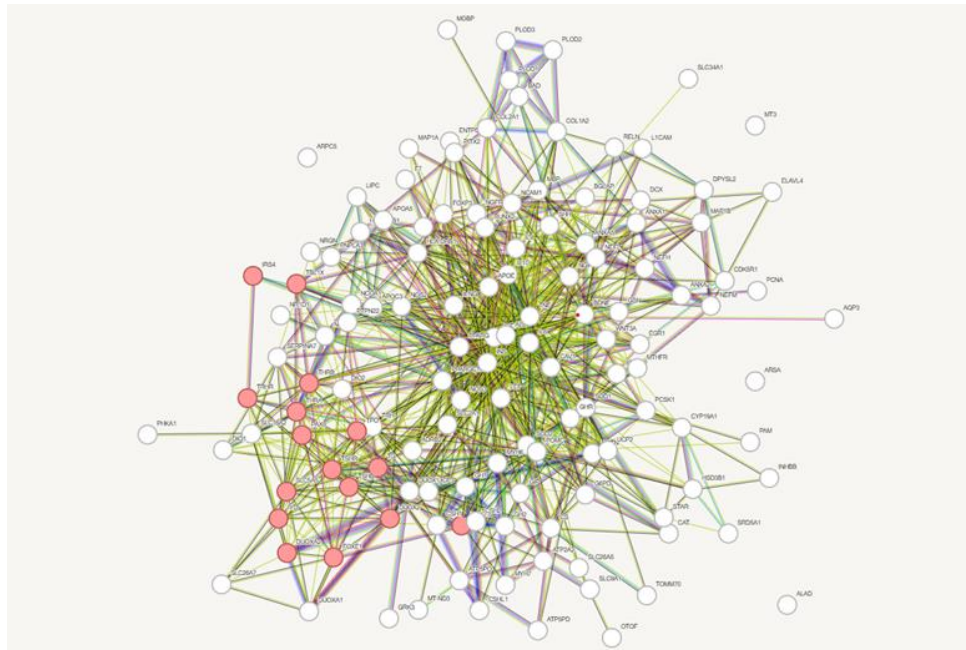


Fig 5.11: Network Showing the Highlighted Genes Involved in Hypothyroidism

5.2.3 NETWORK SHOWING THE HIGHLIGHTED GENES INVOLVED IN THE PATHWAY OF HYPOTHYROIDISM

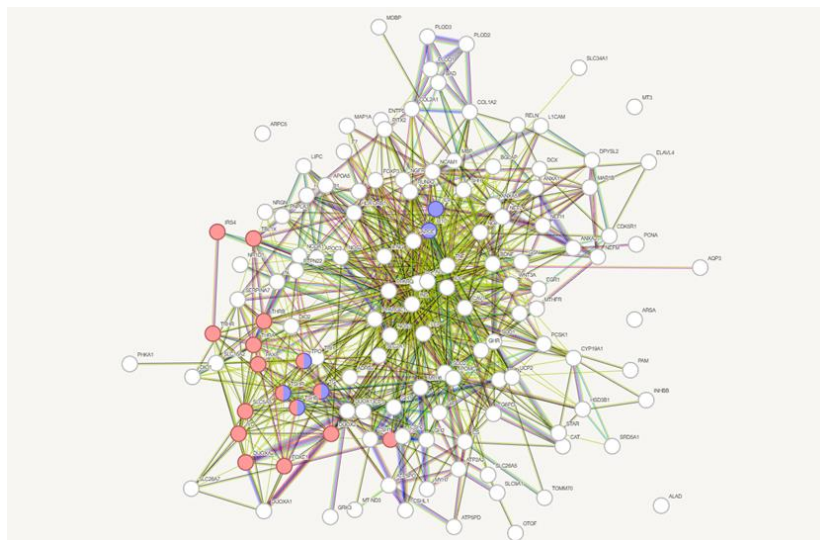


Fig 5.12: Network Showing The Highlighted Genes Involved In The Pathway of hypothyroidism

5.3 CYTOSCAPE

The network was loaded, and basic functions of interactions and node and edge editing, and the layout of the network were modified (Franz et al., 2023; Nishida et al., 2014). Visualised network from the string database was exported to Cytoscape for visualization and identification of the sub networks and hubs present in the network using Mcode and Cytohubba, plug-in present in Cytoscape.

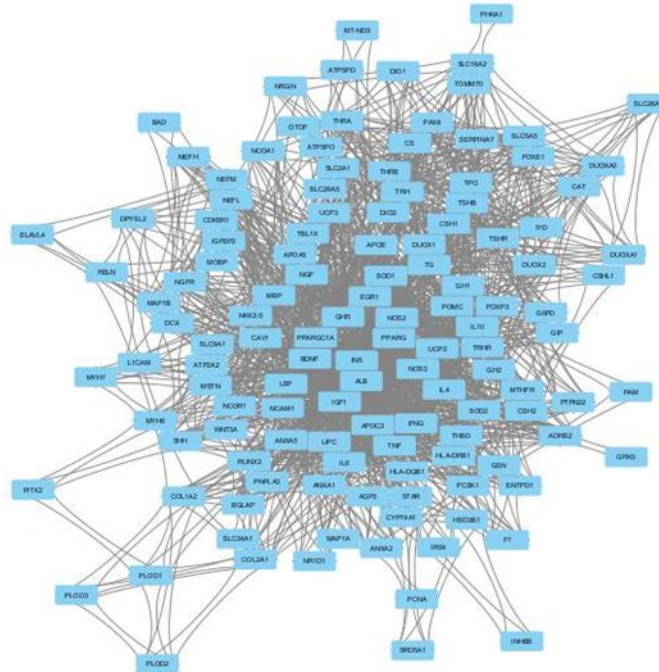


Figure 5.13 PPI - Network using Cytoscape

Summary Statistics
Number of nodes: 129
Number of edges: 1920
Avg. number of neighbors: 14.884
Network diameter: 5
Network radius: 3
Characteristic path length: 2.290
Clustering coefficient: 0.562
Network density: 0.116
Network heterogeneity: 0.879
Network centralization: 0.469
Connected components: 1
Analysis time (sec): 0.074

Figure 5.14 Stats using Analyzer

Genes	Betweenness Centrality	Closeness Centrality	Degree	Number Of Undirected Edges
INS	0.156982054	0.691891892	148	148
ALB	0.125817018	0.666666667	132	132
TNF	0.069561056	0.62745098	110	110
IGF1	0.057296753	0.603773585	96	96
IL6	0.056824743	0.621359223	108	108
SOD1	0.044357626	0.526748971	60	60
BDNF	0.038952301	0.558951965	74	74
PPARG	0.036314381	0.592592593	92	92
APOE	0.035864046	0.531120332	70	70
POMC	0.035717775	0.537815126	68	68
TG	0.03396702	0.526748971	66	66
LEP	0.031132162	0.563876652	84	84
TSHR	0.029355505	0.503937008	60	60
NGF	0.02762835	0.52892562	56	56
MBP	0.022750661	0.486692015	42	42
PPARGC1A	0.022524335	0.533333333	62	62
CYP19A1	0.02102217	0.468864469	26	26
COL2A1	0.019654815	0.458781362	26	26
SOD2	0.019021495	0.520325203	54	54
COL1A2	0.018082186	0.455516014	26	26

Fig 5.14: Top genes identified using Analyzer

5.3.1 MCODE-Identification of subnetwork

I started by building a protein-protein interaction network in the STRING database. This network was then exported to Cytoscape for in-depth visualization (Nishida et al., 2014). Using Cytoscape's MCODE plugin, I searched for highly connected subnetworks (clusters) within the larger network. MCODE successfully identified two significant clusters:

Cluster 1:

24 nodes (proteins)

382 edges (interactions between the proteins)

MCODE score: 16.609

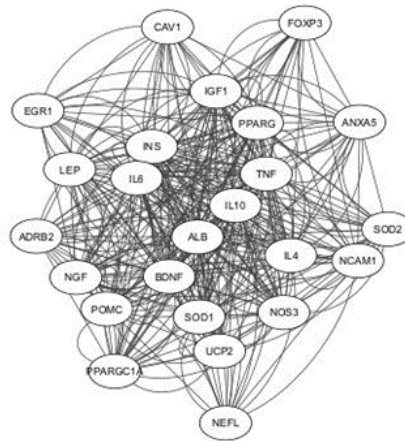


Fig 5.15: MCODE Subnetwork - Cluster 1

Genes	Betweenness Centrality	Closeness Centrality	Degree	MCODE::Score (1)	Number Of Undirected Edges
INS	0.156982054	0.691891892	148	9.836666667	148
ALB	0.125817018	0.666666667	132	9.836666667	132
TNF	0.069561056	0.62745098	110	9.836666667	110
IGF1	0.057296753	0.603773585	96	11.89542484	96
IL6	0.056824743	0.621359223	108	9.836666667	108
SOD1	0.044357626	0.526748971	60	12	60
BDNF	0.038952301	0.558951965	74	11.89542484	74
PPARG	0.036314381	0.592592593	92	9.836666667	92
POMC	0.035717775	0.537815126	68	12	68
LEP	0.031132162	0.563876652	84	12.13970588	84
NGF	0.02762835	0.52892562	56	11.86813187	56
PPARGC1A	0.022524335	0.533333333	62	11.54285714	62
SOD2	0.019021495	0.520325203	54	12	54
ADRB2	0.01637664	0.492307692	34	10.63736264	34
ANXA5	0.015708316	0.505928854	48	11	48
IL10	0.011763889	0.537815126	70	11.63157895	70
NCAM1	0.01000912	0.507936508	46	9.848484848	46
NEFL	0.006582402	0.47761194	36	10	36
IL4	0.005554583	0.507936508	54	9.978947368	54
EGR1	0.004025085	0.486692015	30	11	30
CAV1	0.003494917	0.501960784	42	11	42
NOS3	0.002628464	0.507936508	46	10.10526316	46
FOXP3	0.001494688	0.479400749	38	11	38
UCP2	0.001216107	0.484848485	34	12	34

Fig 5.16: Genes/Proteins associated in Cluster 1

Key Players in Hypothyroidism: The high MCODE scores highlight these genes as potentially central to the disease process.

Diverse Functions: The genes have various functions like metabolism (INS, IGF1), inflammation (TNF, IL6), and cell signaling (BDNF, NGF). This suggests that hypothyroidism might involve disruptions in multiple biological pathways.

Potential Therapeutic Targets: Understanding the interactions within this cluster could reveal targets for drug interventions to treat hypothyroidism.

Cluster 2:

27 nodes (proteins)

166 edges (interactions)

MCODE score: 6.385

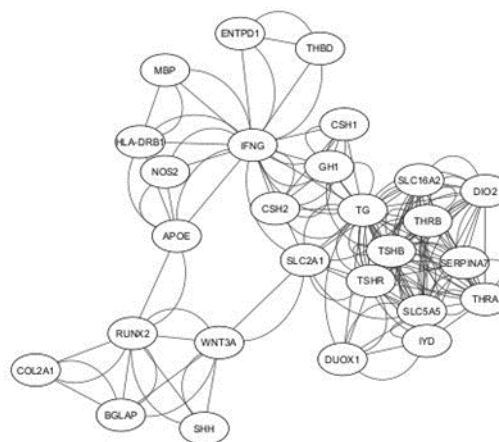


Fig 5.17: MCODE Subnetwork - Cluster 2

Genes	Betweenness Centrality	Closeness Centrality	Degree	MCODE::Score (1)	Number Of Undirected Edges
APOE	0.035864046	0.531120332	70	9.391304348	70
TG	0.03396702	0.526748971	66	8.791208791	66
TSHR	0.029355505	0.503937008	60	8.791208791	60
MBP	0.022750661	0.486692015	42	9.340659341	42
COL2A1	0.019654815	0.458781362	26	8	26
SLC16A2	0.017334002	0.41025641	32	9	32
IFNG	0.016209733	0.554112554	74	8.566153846	74
BGLAP	0.016057469	0.463768116	28	8.672727273	28
THRA	0.013016659	0.457142857	40	9	40
THRB	0.011577026	0.488549618	42	8.076923077	42
SLC2A1	0.011464556	0.516129032	44	9.487179487	44
SHH	0.01128238	0.47761194	30	7.822222222	30
DUOX1	0.008918803	0.488549618	32	8	32
DIO2	0.008778725	0.486692015	38	9	38
TSHB	0.006689995	0.449122807	42	8.791208791	42
SERPINA7	0.004006136	0.460431655	28	9	28
IYD	0.003642941	0.414239482	36	7.846153846	36
SLC5A5	0.003581706	0.414239482	40	8.791208791	40
GH1	0.00280548	0.479400749	30	8.076923077	30
RUNX2	0.001834783	0.47761194	32	8.454545455	32
CSH1	0.001116482	0.465454545	24	8.318181818	24
CSH2	0.001116482	0.465454545	24	8.318181818	24
WNT3A	9.18E-04	0.467153285	24	8	24
HLA-DRB1	8.19E-04	0.462093863	26	7.822222222	26
NOS2	3.57E-04	0.479400749	28	9.615384615	28
THBD	2.61E-04	0.43537415	24	9	24
ENTPD1	0	0.430976431	18	9	18

Fig 5.18: Genes/Proteins associated with Cluster2

These genes and their high MCODE scores suggest that they form a tightly connected subnetwork within the larger protein-protein interaction network. This subnetwork likely plays an important role in biological processes relevant to hypothyroidism.

5.3.2 CLUEGO – FUNCTIONAL ENRICHMENT ANALYSIS OF SUBNETWORKS

5.3.2.1 BIOLOGICAL PROCESS OF CLUSTER 1

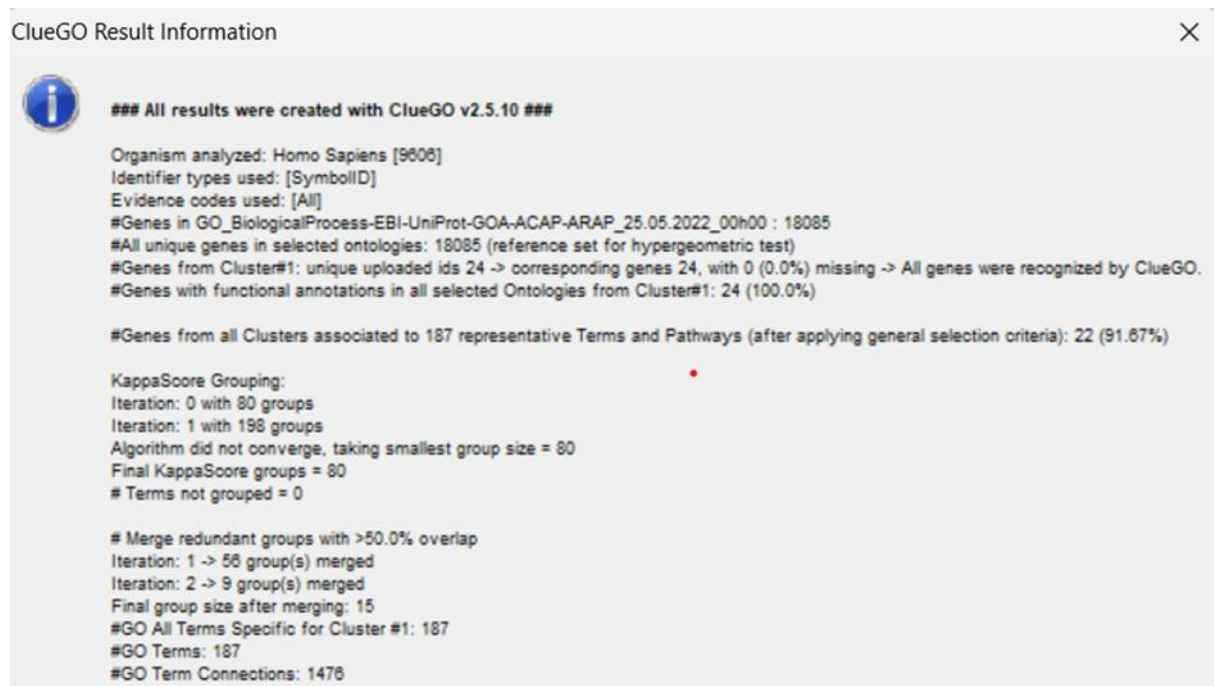


Fig 5.19: Result Of ClueGO plugin

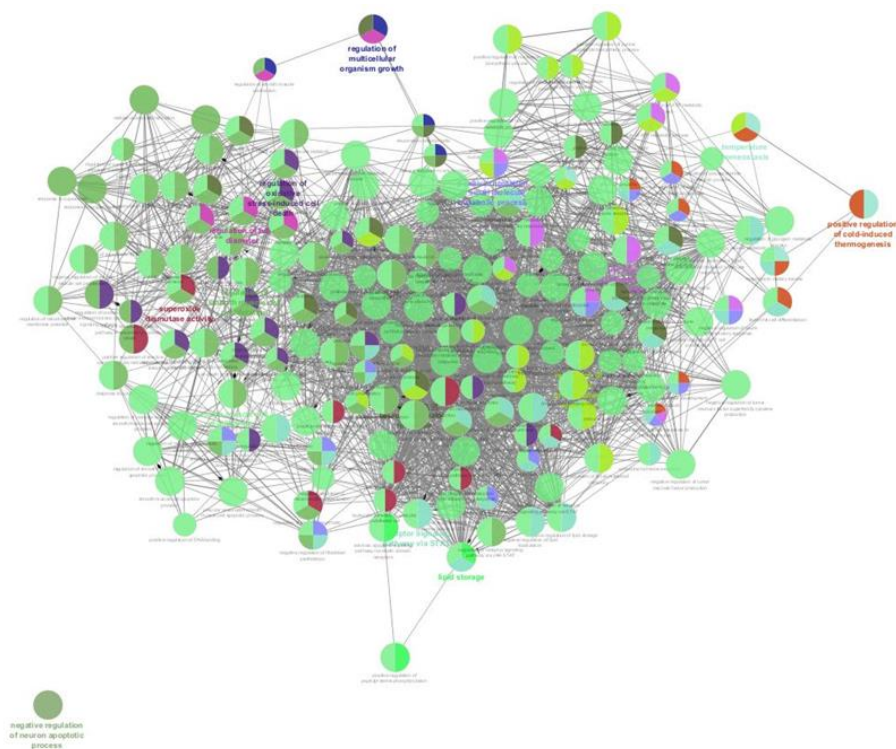


Fig 5.20: Network Highlights Biological Process In Cluster 1

5.3.2.2 KEGG PATHWAY OF CLUSTER 1

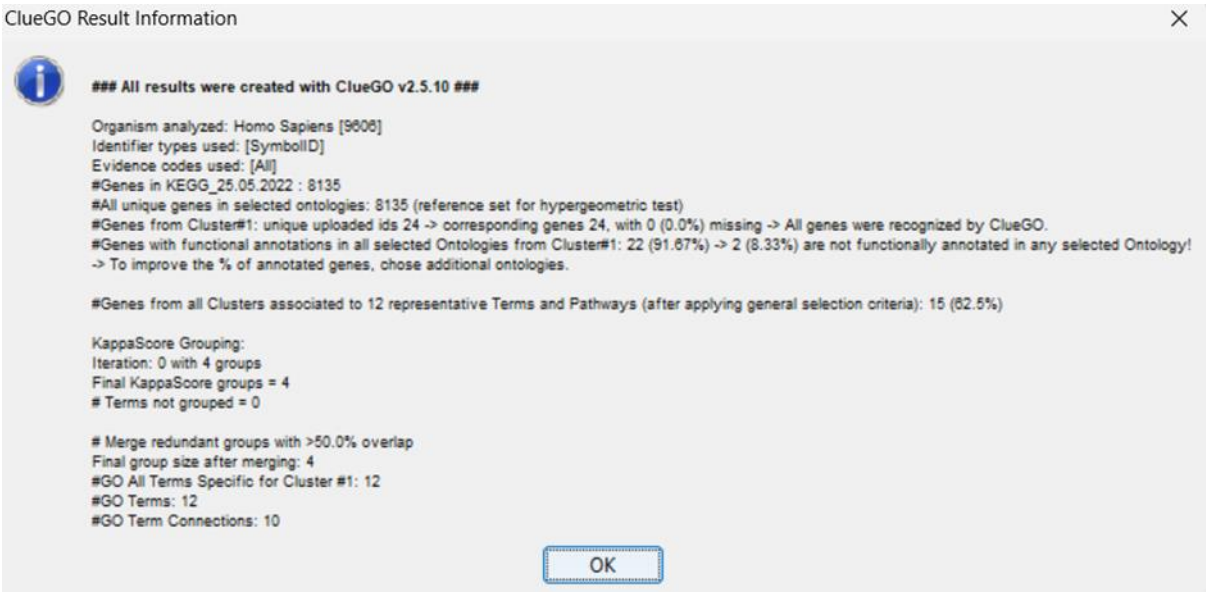


Fig 5.21: Result of KEGG Pathway using ClueGO

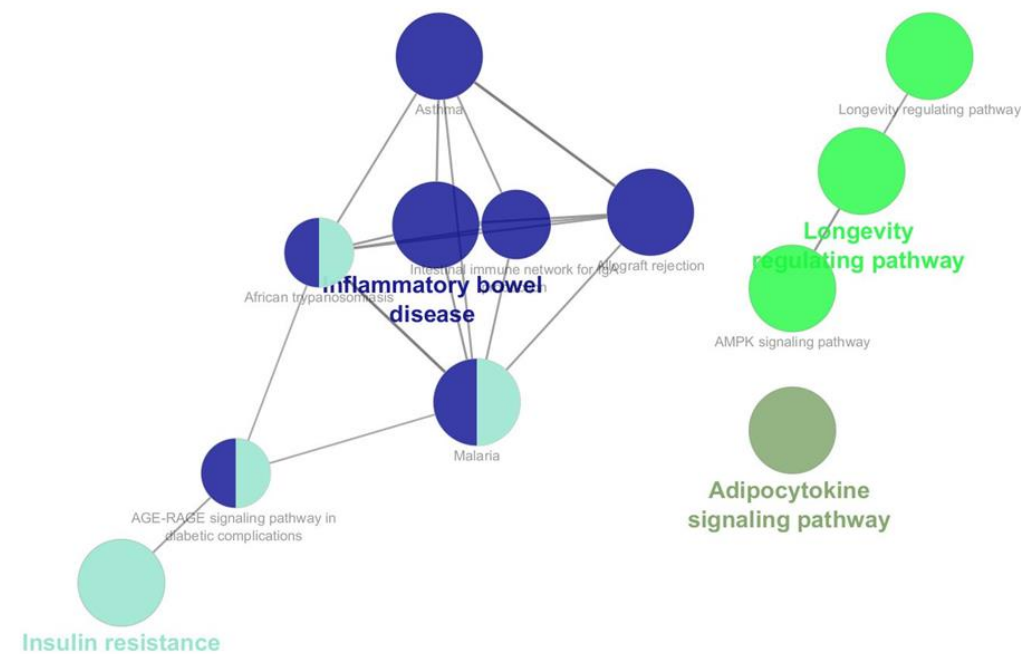


Fig 5.22: Network Highlights KEGG Pathways In Cluster 1

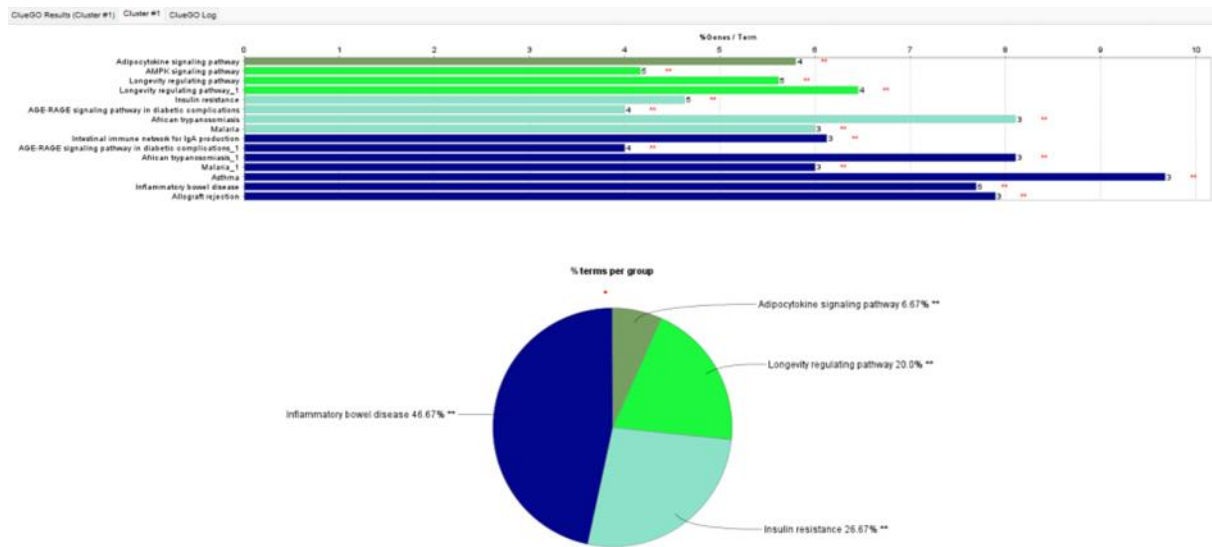


Fig 5.23: Result of KEGG Pathway Analysis using ClueGO

5.3.2.3 BIOLOGICAL PROCESS OF CLUSTER 2

ClueGO Result Information

All results were created with ClueGO v2.5.10

Organism analyzed: Homo Sapiens [9606]
 Identifier types used: [SymbolID]
 Evidence codes used: [All]
 #Genes in GO_BiologicalProcess-EBI-UniProt-GOA-ACAP-ARAP_25.05.2022_00h00 : 18085
 #All unique genes in selected ontologies: 18085 (reference set for hypergeometric test)
 #Genes from Cluster#1: unique uploaded ids 27 -> corresponding genes 27, with 0 (0.0%) missing -> All genes were recognized by ClueGO.
 #Genes with functional annotations in all selected Ontologies from Cluster#1: 27 (100.0%)

#Genes from all Clusters associated to 21 representative Terms and Pathways (after applying general selection criteria): 21 (77.78%)

KappaScore Grouping:
 Iteration: 0 with 7 groups
 Final KappaScore groups = 7
 # Terms not grouped = 0

Merge redundant groups with >50.0% overlap
 Final group size after merging: 7
 #GO All Terms Specific for Cluster #1: 21
 #GO Terms: 21
 #GO Term Connections: 25

OK

Fig 5.24: Result Of ClueGO plugin for Cluster 2

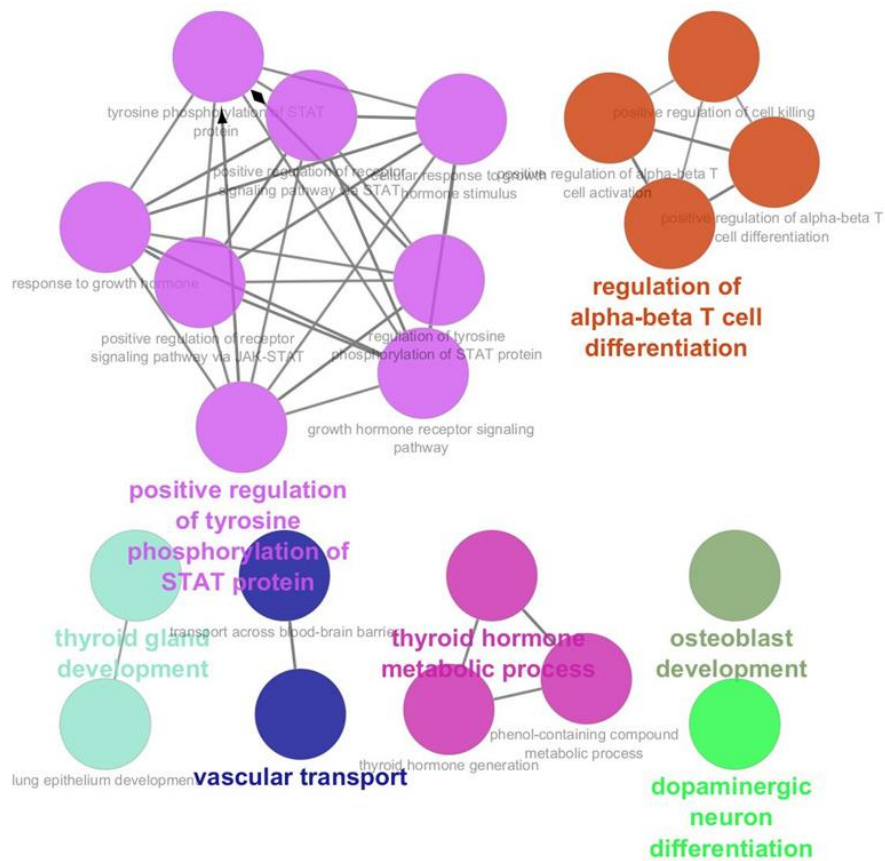


Fig 5.25: Network Highlights Biological Process In Cluster 2

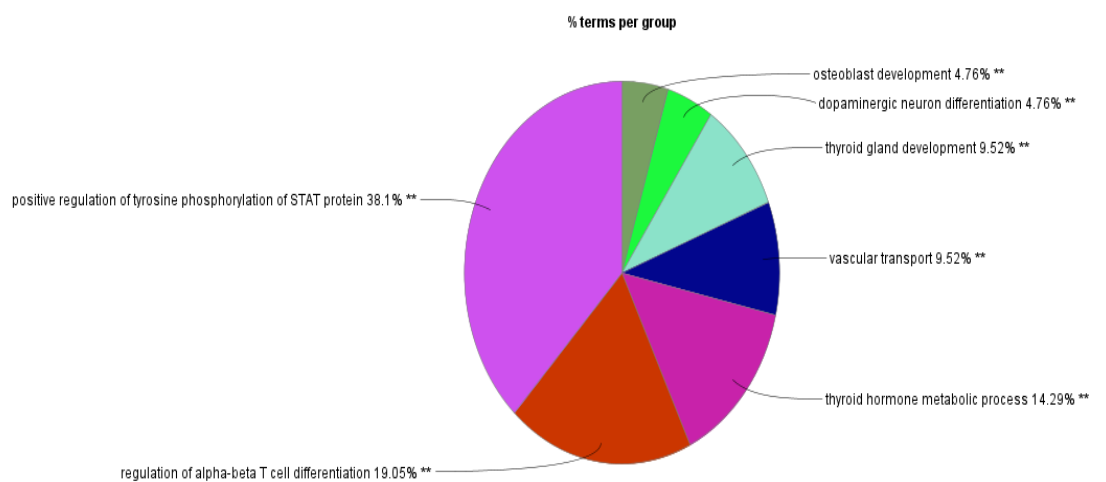


Fig 5.26: Result of Biological Process For Cluster 2

5.3.2.4 KEGG PATHWAY OF CLUSTER 2

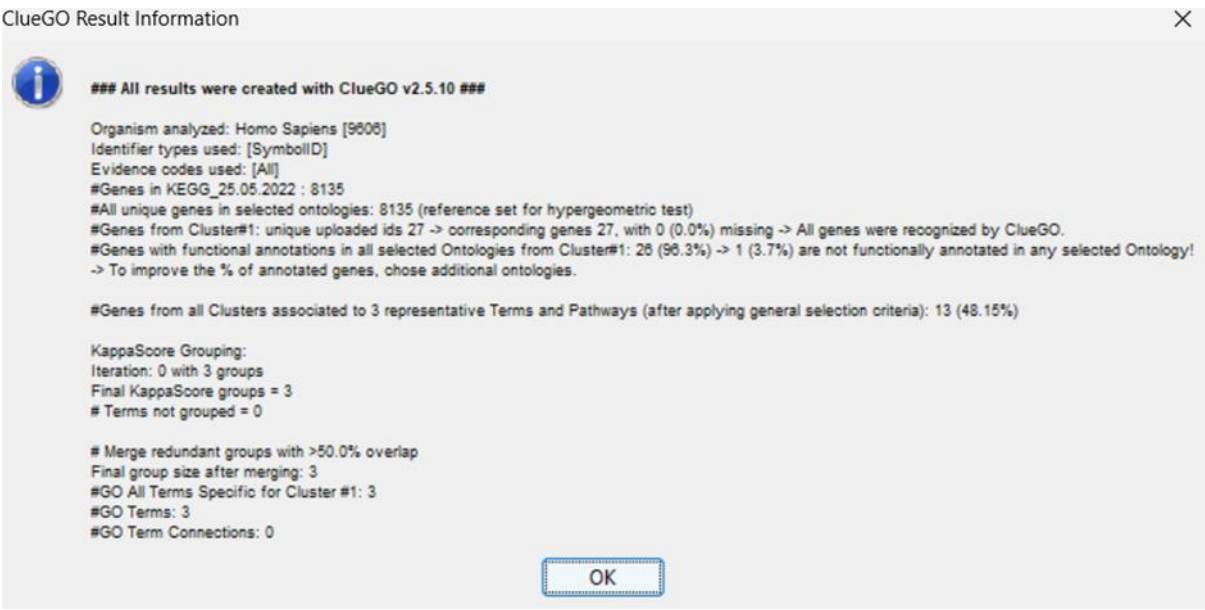


Fig 5.27: Result of KEGG Pathway using ClueGO

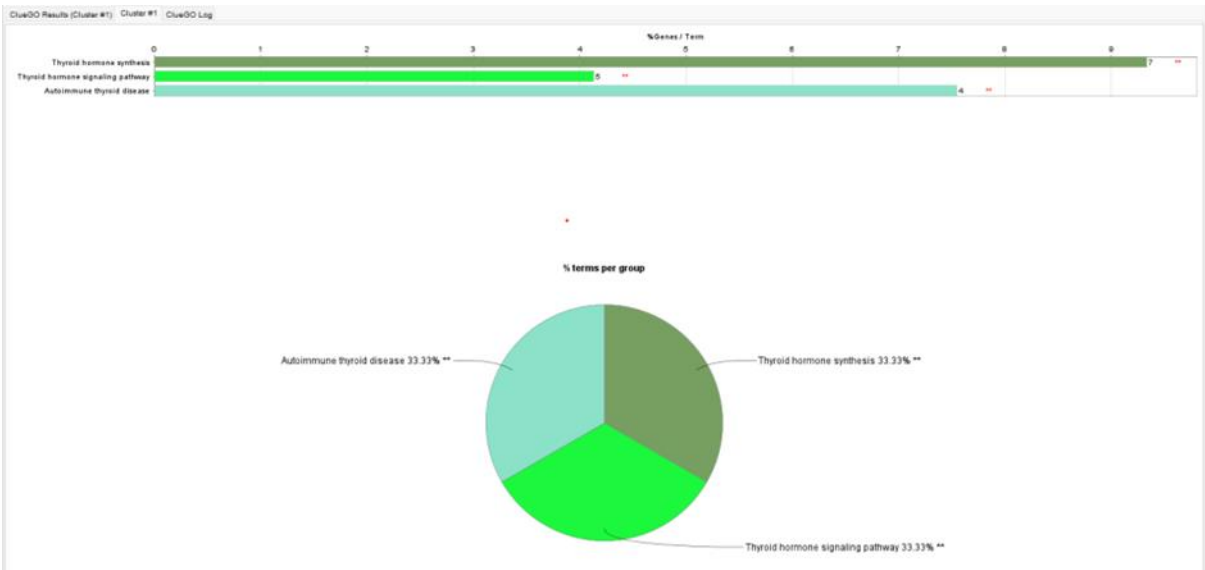


Fig 5.28: Result of KEGG Pathway analysis For Cluster 2

RESULT:

Both clusters likely play a role in the complex picture of hypothyroidism but in different ways:

Cluster 1: May represent the broader systemic effects of hypothyroidism or underlying inflammatory processes contributing to the disease.

Cluster 2: Likely reflects the disruption of thyroid hormone production and signaling pathways, which is central to hypothyroidism.

5.3.3 CYTOHUBBA – IDENTIFICATION OF HUB GENES

Genes having more interactions with other genes are defined as hub genes. Genes with greater degree scores have more interactions with other genes.

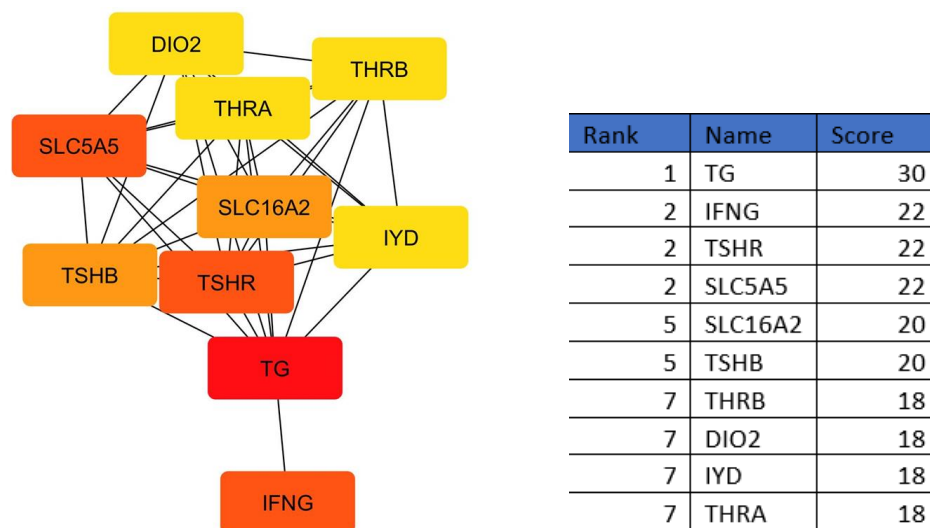


Fig 5.29: HubGenes and score

- TG (Thyroglobulin): This is the primary protein in the thyroid gland. It's the building block for thyroid hormones (T3 and T4).
- IFNG (Interferon Gamma): A cytokine (immune signaling molecule) that can play a role in autoimmune thyroid disorders like Hashimoto's thyroiditis that can cause hypothyroidism.

- TSHR (Thyroid Stimulating Hormone Receptor): The receptor on thyroid cells that binds to TSH (Thyroid Stimulating Hormone), the signal from the pituitary gland that tells the thyroid to make hormones.
- SLC5A5 (Sodium/Iodide Symporter): Crucial for transporting iodine into thyroid cells, which is essential for making thyroid hormone.
- SLC16A2 (MCT8): A transporter responsible for moving thyroid hormones into and out of cells. It plays a significant role in how thyroid hormones work throughout the body.
- TSHB (Thyroid Stimulating Hormone Beta Subunit): One part of the TSH complex, the hormone that stimulates the thyroid gland.
- THRB (Thyroid Hormone Receptor Beta): The receptor for thyroid hormones. Thyroid hormones bind to this receptor to regulate gene expression and cellular function.
- DIO2 (Type 2 Deiodinase): This enzyme converts the less active thyroid hormone (T4) into the more active form (T3).
- IYD (Iodotyrosine Deiodinase): Involved in recycling iodine within the thyroid gland for the efficient production of thyroid hormones.
- THRA (Thyroid Hormone Receptor Alpha): Another type of thyroid hormone receptor, playing a role in regulating metabolism and development.

Causes of Hypothyroidism:

- Mutations in TG could lead to poor thyroid hormone production.
- IFNG's presence suggests a potential autoimmune component.
- Problems with SLC5A5 might indicate iodine deficiency.

LIST OF TABLES

Table 1: Top Genes/Proteins Based On Degree

Genes	Betweenness Centrality	Closeness Centrality	Degree	Number Of Undirected Edges
INS	0.156982054	0.691891892	148	148
ALB	0.125817018	0.666666667	132	132
TNF	0.069561056	0.62745098	110	110
IGF1	0.057296753	0.603773585	96	96
IL6	0.056824743	0.621359223	108	108
SOD1	0.044357626	0.526748971	60	60
BDNF	0.038952301	0.558951965	74	74
PPARG	0.036314381	0.592592593	92	92
APOE	0.035864046	0.531120332	70	70
POMC	0.035717775	0.537815126	68	68
TG	0.03396702	0.526748971	66	66
LEP	0.031132162	0.563876652	84	84
TSHR	0.029355505	0.503937008	60	60
NGF	0.02762835	0.52892562	56	56
MBP	0.022750661	0.486692015	42	42
PPARGC1A	0.022524335	0.533333333	62	62
CYP19A1	0.02102217	0.468864469	26	26
COL2A1	0.019654815	0.458781362	26	26
SOD2	0.019021495	0.520325203	54	54
COL1A2	0.018082186	0.455516014	26	26

Table 2 : MCODE Subnetwork - Cluster 1

Genes	Betweenness Centrality	Closeness Centrality	Degree	MCODE :: Score (1)	Number Of Undirected Edges
INS	0.156982054	0.691891892	148	9.836666667	148
ALB	0.125817018	0.666666667	132	9.836666667	132
TNF	0.069561056	0.62745098	110	9.836666667	110
IGF1	0.057296753	0.603773585	96	11.89542484	96
IL6	0.056824743	0.621359223	108	9.836666667	108
SOD1	0.044357626	0.526748971	60	12	60
BDNF	0.038952301	0.558951965	74	11.89542484	74
PPARG	0.036314381	0.592592593	92	9.836666667	92
POMC	0.035717775	0.537815126	68	12	68
LEP	0.031132162	0.563876652	84	12.13970588	84
NGF	0.02762835	0.52892562	56	11.86813187	56
PPARGC1A	0.022524335	0.533333333	62	11.54285714	62
SOD2	0.019021495	0.520325203	54	12	54
ADRB2	0.01637664	0.492307692	34	10.63736264	34
ANXA5	0.015708316	0.505928854	48	11	48
IL10	0.011763889	0.537815126	70	11.63157895	70
NCAM1	0.010000912	0.507936508	46	9.848484848	46
NEFL	0.006582402	0.47761194	36	10	36
IL4	0.005554583	0.507936508	54	9.978947368	54
EGR1	0.004025085	0.486692015	30	11	30
CAV1	0.003494917	0.501960784	42	11	42
NOS3	0.002628464	0.507936508	46	10.10526316	46
FOXP3	0.001494688	0.479400749	38	11	38
UCP2	0.001216107	0.484848485	34	12	34

Table 3: MCODE Subnetwork - Cluster 2

Genes	Betweenness Centrality	Closeness Centrality	Degree	MCODE::Score (1)	Number Of Undirected Edges
APOE	0.035864046	0.531120332	70	9.391304348	70
TG	0.03396702	0.526748971	66	8.791208791	66
TSHR	0.029355505	0.503937008	60	8.791208791	60
MBP	0.022750661	0.486692015	42	9.340659341	42
COL2A1	0.019654815	0.458781362	26	8	26
SLC16A2	0.017334002	0.41025641	32	9	32
IFNG	0.016209733	0.554112554	74	8.566153846	74
BGLAP	0.016057469	0.463768116	28	8.672727273	28
THRA	0.013016659	0.457142857	40	9	40
THRB	0.011577026	0.488549618	42	8.076923077	42
SLC2A1	0.011464556	0.516129032	44	9.487179487	44
SHH	0.01128238	0.47761194	30	7.822222222	30
DUOX1	0.008918803	0.488549618	32	8	32
DIO2	0.008778725	0.486692015	38	9	38
TSHB	0.006689995	0.449122807	42	8.791208791	42
SERPINA7	0.004006136	0.460431655	28	9	28
IYD	0.003642941	0.414239482	36	7.846153846	36
SLC5A5	0.003581706	0.414239482	40	8.791208791	40
GH1	0.00280548	0.479400749	30	8.076923077	30
RUNX2	0.001834783	0.47761194	32	8.454545455	32
CSH1	0.001116482	0.465454545	24	8.318181818	24
CSH2	0.001116482	0.465454545	24	8.318181818	24
WNT3A	9.18E-04	0.467153285	24	8	24
HLA-DRB1	8.19E-04	0.462093863	26	7.822222222	26
NOS2	3.57E-04	0.479400749	28	9.615384615	28
THBD	2.61E-04	0.43537415	24	9	24
ENTPD1	0	0.430976431	18	9	18

Table 4: HubGenes

Rank	Name	Score
1	TG	30
2	IFNG	22
2	TSHR	22
2	SLC5A5	22
5	SLC16A2	20
5	TSHB	20
7	THRB	18
7	DIO2	18
7	IYD	18
7	THRA	18

CHAPTER 6

SUMMARY AND CONCLUSION

6.1 SUMMARY

This bioinformatics project aimed to illuminate the genetic underpinnings of hypothyroidism. Starting with a refined list of 152 genes, I employed the DAVID tool to conduct functional enrichment analysis. Key biological processes identified included thyroid hormone biosynthesis, oxidoreductase activity, and hormone activity. The KEGG pathway analysis revealed significant associations with thyroid hormone signaling, thyroid hormone synthesis, and autoimmune thyroid disease. Enrichment analysis of cellular components highlighted the "secreted" category, aligning with the thyroid gland's role in hormone production.

To explore gene interactions, I constructed a protein-protein interaction (PPI) network using STRING. The network's high PPI enrichment p-value ($<1.0e-16$) suggests a strong biological interconnectedness among the proteins involved. The PPI network was transferred to Cytoscape for further analysis. Using the MCODE plugin, I identified two significant clusters of densely connected genes. Cluster 1 (24 nodes, 382 edges) included genes involved in diverse functions like metabolism, inflammation, and cell signaling. Cluster 2 (27 nodes, 166 edges) was enriched in pathways related to thyroid hormone production and signaling. Further analysis with ClueGO confirmed the biological processes and pathways associated with both clusters, underscoring their potential relevance to hypothyroidism.

Finally, I used the CytoHubba plugin to pinpoint hub genes within the PPI network. The top hub genes identified were TG, IFNG, TSHR, SLC5A5, SLC16A2, TSHB, THRB, DIO2, IYD, and THRA. Based on the overall analysis, Cluster 2, along with its hub genes, appears to be most directly linked to the disruption of thyroid hormone pathways central to hypothyroidism.

These findings offer valuable insights into the complex genetic landscape of hypothyroidism, pinpointing key genes, pathways, and potential therapeutic targets. Future work should focus on experimental validation of these findings and investigation of the identified genes and pathways as possible targets for new diagnostic tools and precision-based treatments for hypothyroidism.

6.2 CONCLUSION

This bioinformatics project successfully identified key genes and pathways potentially involved in the pathogenesis of hypothyroidism. The analysis revealed significant dysregulation in genes related to thyroid hormone production and immune response. Hub genes, such as [TG, IFNG, TSHR, SLC5A5, SLC16A2, TSHB, THRB, DIO2, IYD, and THRA], were found to be central within the PPI network. Further investigation of these targets could lead to the development of new diagnostic approaches and targeted therapies for hypothyroidism. Future studies should focus on experimental validation and exploring the therapeutic potential of the identified pathways.

REFERENCES

- Bader, G. D., & Hogue, C. W. V. (2003). An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics*, 4, 2.
- Bindea, G., Mlecnik, B., Hackl, H., Charoentong, P., Tosolini, M., Kirilovsky, A., Fridman, W.-H., Pagès, F., Trajanoski, Z., & Galon, J. (2009). ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics*, 25(8), 1091–1093.
- Chaker, L., Razvi, S., Bensenor, I. M., Azizi, F., Pearce, E. N., & Peeters, R. P. (2022). Hypothyroidism. *Nature Reviews. Disease Primers*, 8(1), 30.
- Chin, C.-H., Chen, S.-H., Wu, H.-H., Ho, C.-W., Ko, M.-T., & Lin, C.-Y. (2014). cytoHubba: identifying hub objects and sub-networks from complex interactome. *BMC Systems Biology*, 8 Suppl 4(Suppl 4), S11.
- Chiovato, L., Magri, F., & Carlé, A. (2019). Hypothyroidism in Context: Where We've Been and Where We're Going. *Advances in Therapy*, 36(Suppl 2), 47–58.
- Dai, D. Y., & Sternberg, R. J. (2004). *Motivation, Emotion, and Cognition: Integrative Perspectives on Intellectual Functioning and Development*. Routledge.
- Doncheva, N. T., Morris, J. H., Gorodkin, J., & Jensen, L. J. (2019). Cytoscape StringApp: Network Analysis and Visualization of Proteomics Data. *Journal of Proteome Research*, 18(2), 623–632.
- Doncheva, N. T., Morris, J. H., Holze, H., Kirsch, R., Nastou, K. C., Cuesta-Astro, Y., Rattei, T., Szklarczyk, D., von Mering, C., & Jensen, L. J. (2023).

Cytoscape stringApp 2.0: Analysis and Visualization of Heterogeneous Biological Networks. *Journal of Proteome Research*, 22(2), 637–646.

Effraimidis, G., & Wiersinga, W. M. (2014). Mechanisms in endocrinology: autoimmune thyroid disease: old and new players. *European Journal of Endocrinology / European Federation of Endocrine Societies*, 170(6), R241–R252.

Franz, M., Lopes, C. T., Fong, D., Kucera, M., Cheung, M., Siper, M. C., Huck, G., Dong, Y., Sumer, O., & Bader, G. D. (2023). Cytoscape.js 2023 update: a graph theory library for visualization and analysis. *Bioinformatics*, 39(1). <https://doi.org/10.1093/bioinformatics/btad031>

Garber, J. R., Cobin, R. H., Gharib, H., Hennessey, J. V., Klein, I., Mechanick, J. I., Pessah-Pollack, R., Singer, P. A., Woeber, K. A., & American Association of Clinical Endocrinologists and American Thyroid Association Taskforce on Hypothyroidism in Adults. (2012). Clinical practice guidelines for hypothyroidism in adults: cosponsored by the American Association of Clinical Endocrinologists and the American Thyroid Association. *Endocrine Practice: Official Journal of the American College of Endocrinology and the American Association of Clinical Endocrinologists*, 18(6), 988–1028.

Hennessey, J. V. (2023). Levothyroxine Monotherapy: What Works Better for the Individual With Hypothyroidism? *Endocrine Practice: Official Journal of the American College of Endocrinology and the American Association of Clinical Endocrinologists*, 29(7), 572–580.

Kishore, R., Arnaboldi, V., Van Slyke, C. E., Chan, J., Nash, R. S., Urbano, J. M., Dolan, M. E., Engel, S. R., Shimoyama, M., Sternberg, P. W., &

- Genome Resources, T. A. O. (2020). Automated generation of gene summaries at the Alliance of Genome Resources. *Database: The Journal of Biological Databases and Curation*, 2020. <https://doi.org/10.1093/database/baaa037>
- Lu, D. K., Guo, Z. C., Zhang, J. J., Yu, X., & Zhang, Z. Y. (2024). Functional Investigation and Two-sample Mendelian Randomization Study of Inguinal Hernia Hub Genes Obtained by Bioinformatics Analysis. *Current Computer-Aided Drug Design*. <https://doi.org/10.2174/0115734099282407240325054745>
- Majeed, A., & Mukhtar, S. (2023). Protein-Protein Interaction Network Exploration Using Cytoscape. *Methods in Molecular Biology*, 2690, 419–427.
- Nishida, K., Ono, K., Kanaya, S., & Takahashi, K. (2014). KEGGscape: a Cytoscape app for pathway data integration. *F1000Research*, 3, 144.
- Piñero, J., Bravo, À., Queralt-Rosinach, N., Gutiérrez-Sacristán, A., Deu-Pons, J., Centeno, E., García-García, J., Sanz, F., & Furlong, L. I. (2017). DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants. *Nucleic Acids Research*, 45(D1), D833–D839.
- Piñero, J., Saüch, J., Sanz, F., & Furlong, L. I. (2021). The DisGeNET cytoscape app: Exploring and visualizing disease genomics data. *Computational and Structural Biotechnology Journal*, 19, 2960–2967.
- Qiu, K., Li, K., Zeng, T., Liao, Y., Min, J., Zhang, N., Peng, M., Kong, W., & Chen, L.-L. (2021). Integrative Analyses of Genes Associated with Hashimoto's Thyroiditis. *Journal of Immunology Research*, 2021, 8263829.

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., & Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research*, 13(11), 2498–2504.

Sherman, B. T., Hao, M., Qiu, J., Jiao, X., Baseler, M. W., Lane, H. C., Imamichi, T., & Chang, W. (2022). DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). *Nucleic Acids Research*, 50(W1), W216–W221.

Shi, W., Zhang, X., Xu, C., Pang, R., Fan, Z., Wan, X., Jiang, Z., Li, H., Li, Z., & Zhang, H. (2022). Identification of Hub Genes and Pathways Associated with Oxidative Stress of Cartilage in Osteonecrosis of Femoral Head Using Bioinformatics Analysis. *Cartilage*, 13(1), 19476035221074000.

Sue, L. Y., & Leung, A. M. (2020). Levothyroxine for the Treatment of Subclinical Hypothyroidism and Cardiovascular Disease. *Frontiers in Endocrinology*, 11, 591588.

Szklarczyk, D., Kirsch, R., Koutrouli, M., Nastou, K., Mehryary, F., Hachilif, R., Gable, A. L., Fang, T., Doncheva, N. T., Pyysalo, S., Bork, P., Jensen, L. J., & von Mering, C. (2023). The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Research*, 51(D1), D638–D646.

Wiersinga, W. M. (2014). Paradigm shifts in thyroid hormone replacement therapies for hypothyroidism. *Nature Reviews. Endocrinology*, 10(3), 164–174.

Zhang, Y., Garofano, F., Wu, X., Schmid, M., Krawitz, P., Essler, M., & Schmidt-Wolf, I. G. H. (2020). Integrative analysis of key candidate genes and signaling pathways in autoimmune thyroid dysfunction related to anti-CTLA-4 therapy by bioinformatics. *Investigational New Drugs*, 38(6), 1717–1729.

Zheng, L., Dou, X., Song, H., Wang, P., Qu, W., & Zheng, X. (2020). Bioinformatics analysis of key genes and pathways in Hashimoto thyroiditis tissues. *Bioscience Reports*, 40(7). <https://doi.org/10.1042/BSR20200759>