24AIM112 - MOLECULAR BIOLOGY AND BASIC CELLULAR PHYSIOLOGY
24AIM115 - ETHICS, INNOVATIVE RESEARCH AND IPR

GENE CO-EXPRESSION NETWORK ANALYSIS

PRESENTED BY TEAM 10:

MINOTI K   - CB.AI.U4AIM24130

NIVEDHITHA S   - CB.AI.U4AIM24134

JAISON JOHN SAMUEL J   - CB.AI.U4AIM24116

SANCHITA B   - CB.AI.U4AIM24141

FACULTY INCHARGE:

MRS. RESHMA SANAL

DR.NEELESH ASHOK

INTRODUCTION

- **Gene co-expression network analysis** - study gene relationships and functions by constructing networks where **nodes** are genes and **edges** are significant co-expression.

- The **extracted modules** from network can reveal insights to **pathways**, **disease mechanisms**, or **key regulatory genes.**

- Used widely in **genomics** to prioritize candidate genes and understand **molecular interactions** underlaying traits or conditions.

- **Dataset** - from **NCBI (Systems biological assessment of immunity to severe and mild COVID-19 infection (GSE152418)).**

- Extracted **green and turquoise modules** from **cluster dendrogram.**
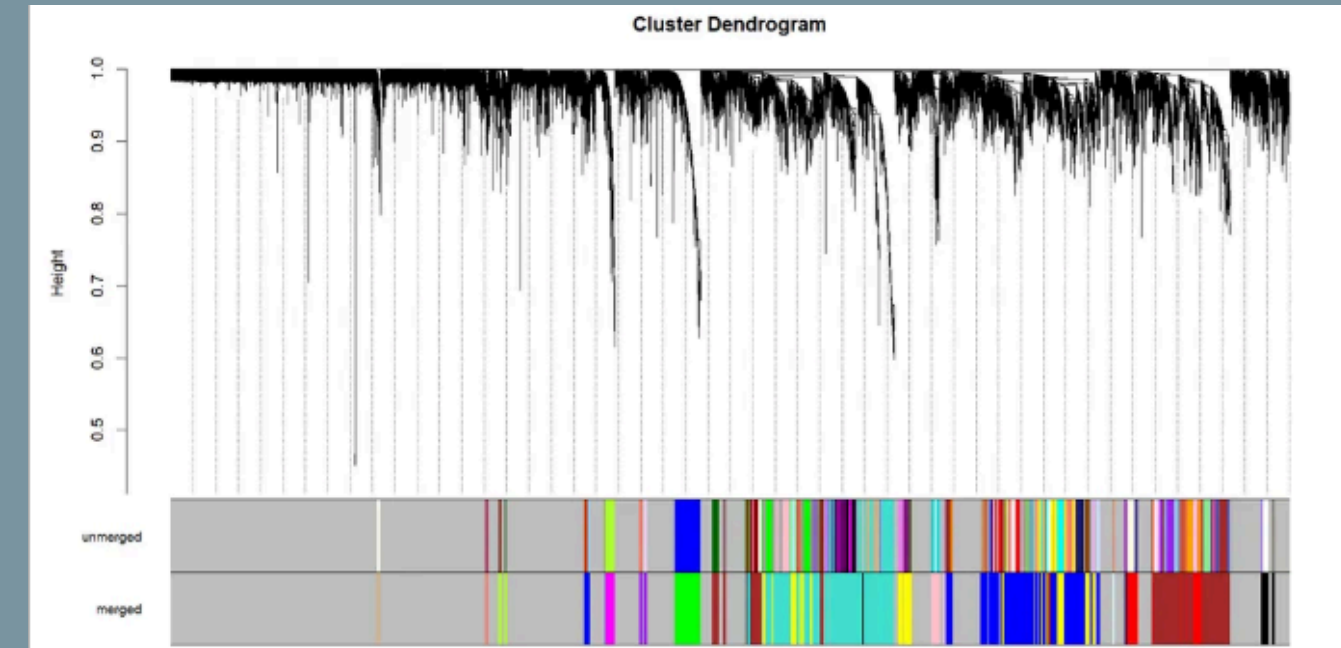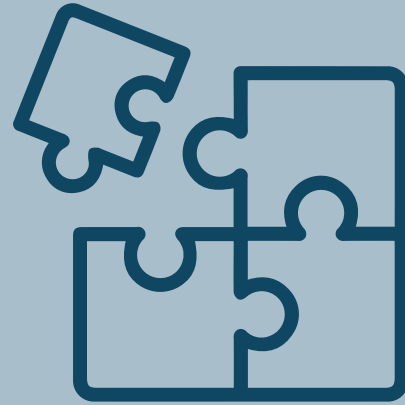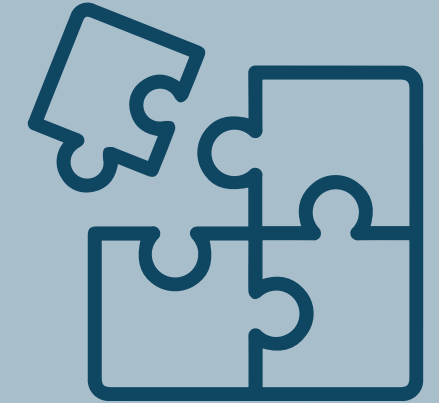

(figure 1. Cluster dendrogram)

OBJECTIVES

To identify gene modules (clusters), hub genes etc., of highly expressed genes with the dataset.

Perform Enrichment analysis of the identified gene modules

Evaluate and understand key information from the above

Analyze the ethical concerns on gene analysis, intellectual property rights and propose solutions for the same.
Explore case studies.

WHAT DID WE DO?

- Performed co-expression network to analyze gene modules, hub genes etc,.
- Plotted cluster dendrogram and collected the inference of the same
- Plotted a heatmap for the dataset and collected the inference for the same
- From the gene modules, extracted the green and turquoise module which had most significant contribution to the disease
- Performed enrichment analysis for the two modules - GO, KEGG
- Plotted enrichment graph and inferred from the same
- Got a broad understanding of working of gene co-expression network analysis
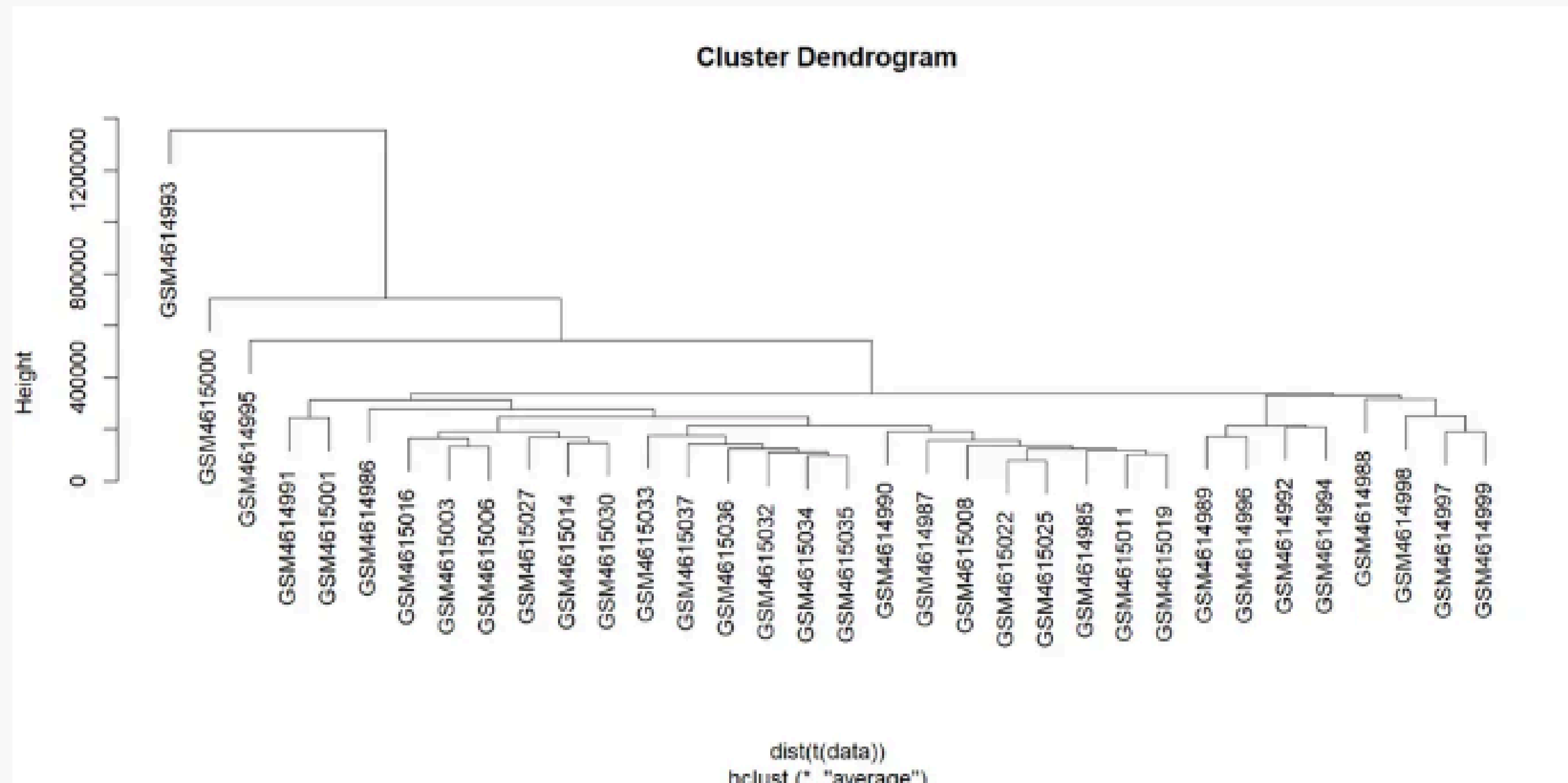
METHODOLOGY

**Co-expression network analysis :-**

- **Expression data** downloaded from NCBI website and metadata (or phenotypic data) taken directly from GEO website
- Both the data are then merged
- Performed quality control - checked for outliers using **goodSamplegenes package, hierarchical clustering, principal component analysis**
- Dendrogram from hierarchical clustering and PCA technique compared and outliers are removed from data
- Created **dds data** with raw counts for **DESeq2** - used to normalize the data
- **WGCNA network** constructed
- A module with **eigengenes** were made to summarize whole module expression and correlated with modules as well
- Binarized the found modules with traits (**disease metadata**)
- Extracted **gene module** and **turquoise module** from the same
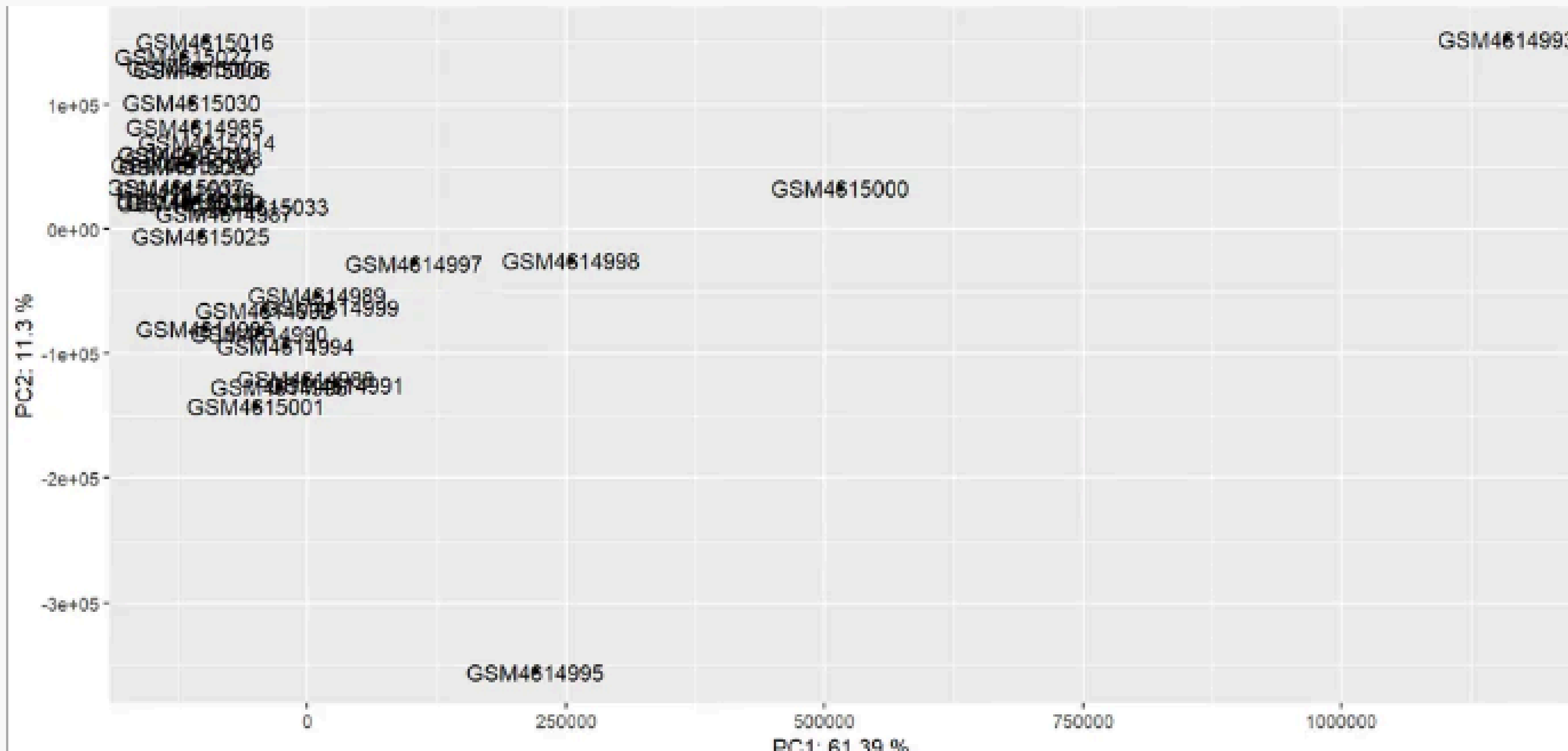
RESULTS

CO-EXPRESSION NETWORK ANALYSIS :-

CLUSTER DENDROGRAM FROM HIERARCHICAL CLUSTERING
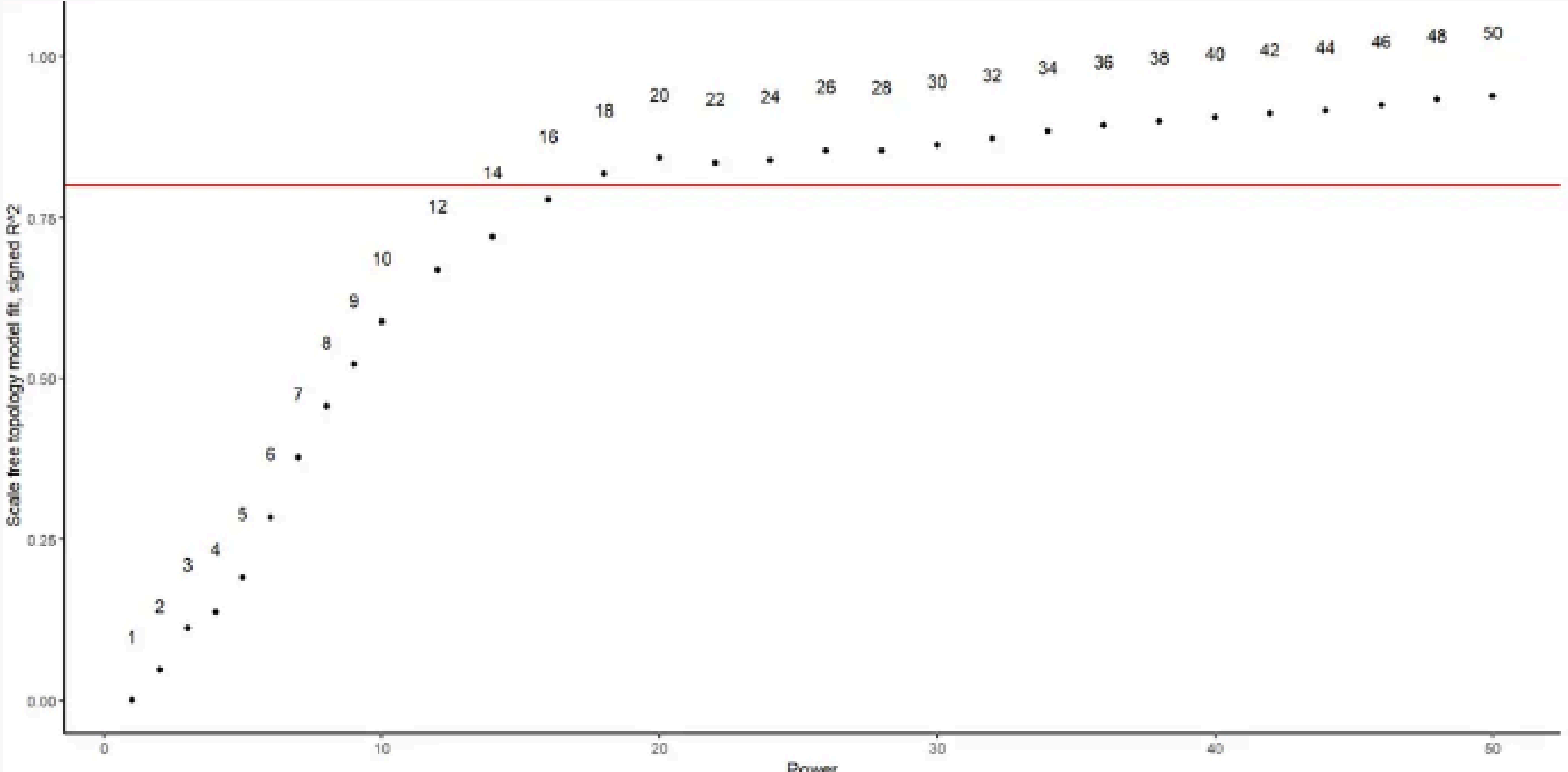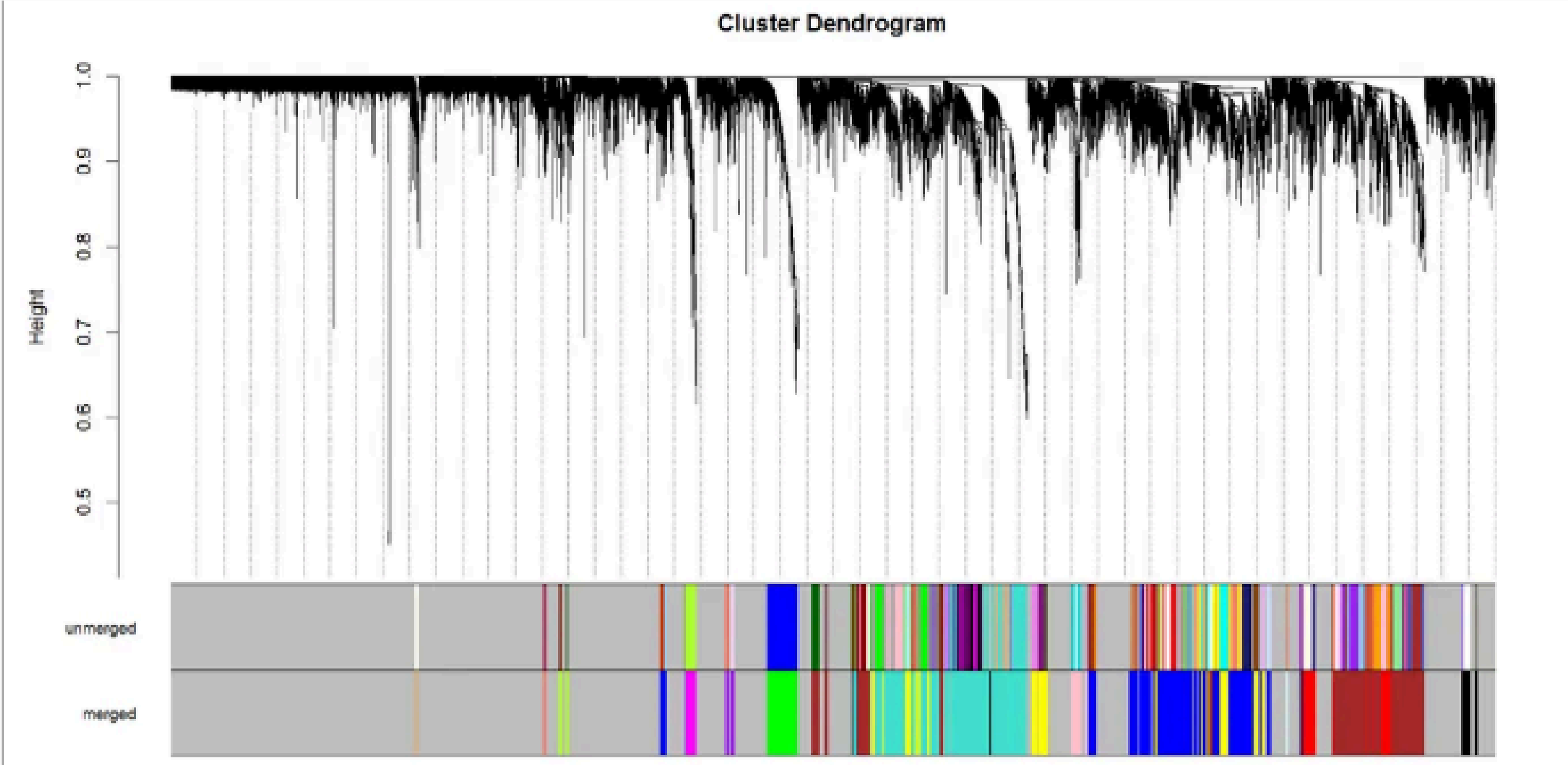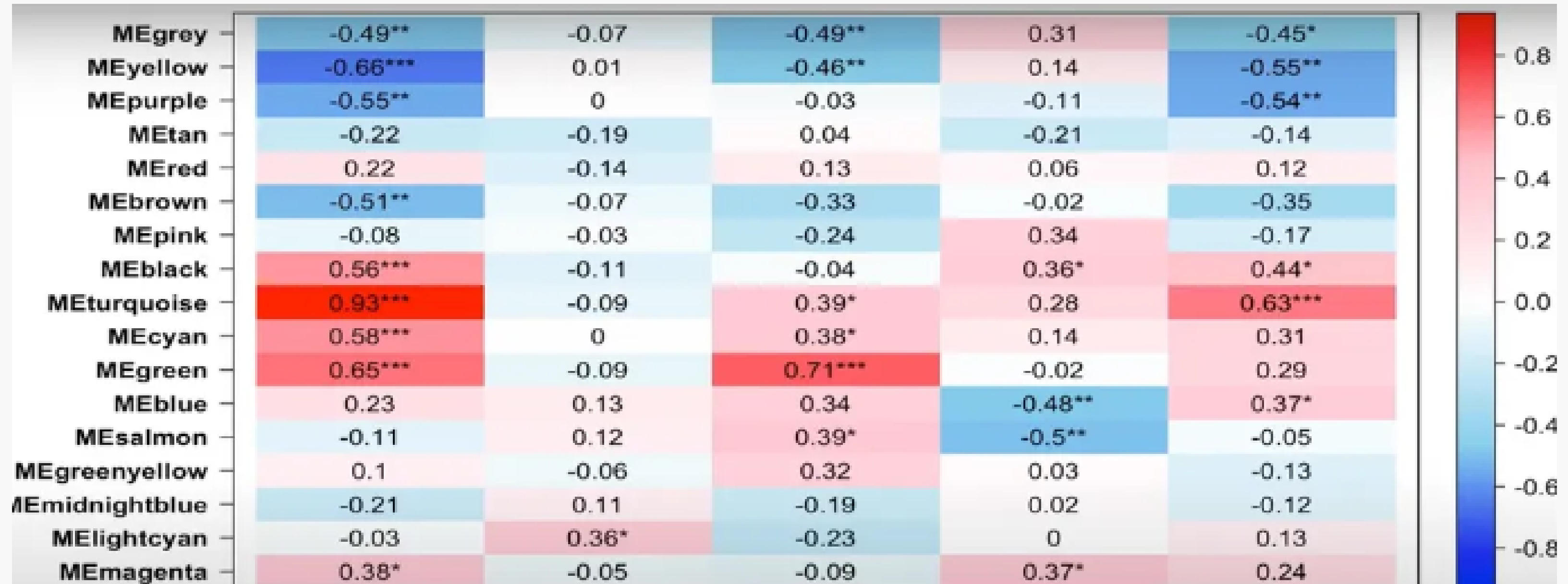
RESULTS

GRAPH USED TO PICK SOFT THRESHOLD

RESULTS

GRAPH FROM PRINCIPLE COMPONENT ANALYSIS

RESULTS

CLUSTER DENDROGRAM

RESULTS

MEgrey	-0.49**	-0.07	-0.49**	0.31	-0.45*
MEyellow	-0.66***	0.01	-0.46**	0.14	-0.55**
MEpurple	-0.55**	0	-0.03	-0.11	-0.54**
MEtan	-0.22	-0.19	0.04	-0.21	-0.14
MEred	0.22	-0.14	0.13	0.06	0.12
MEbrown	-0.51**	-0.07	-0.33	-0.02	-0.35
MEpink	-0.08	-0.03	-0.24	0.34	-0.17
MEblack	0.56***	-0.11	-0.04	0.36*	0.44*
MEturquoise	0.93***	-0.09	0.39*	0.28	0.63***
MEcyan	0.58***	0	0.38*	0.14	0.31
MEgreen	0.65***	-0.09	0.71***	-0.02	0.29
MEblue	0.23	0.13	0.34	-0.48**	0.37*
MEsalmon	-0.11	0.12	0.39*	-0.5**	-0.05
MEgreenyellow	0.1	-0.06	0.32	0.03	-0.13
MEmidnightblue	-0.21	0.11	-0.19	0.02	-0.12
MElightcyan	-0.03	0.36*	-0.23	0	0.13
MEmagenta	0.38*	-0.05	-0.09	0.37*	0.24

HEATMAP FOR GENE MODULES

METHODOLOGY

Enrichment analysis :-

GO enrichment - Analyzes **Biological Process (BP)**, **Molecular Function (MF)**, and **Cellular Component (CC) ontologies**
KEGG pathways - Identifies enriched KEGG pathways
Reactome pathways - examines **reactome database** pathways
WikiPathways - examines from WikiPathways database
Wikipathways- aims to capture a wider range of pathway knowledge, including less well-defined or emerging pathways.

- Converted **Ensembl IDs** to **Entrez IDs**
- Generate visualizations for enrichment results - **Dotplot** (shows top-enriched terms with p-values and gene ratios).
- Created a **summary table** for summarizing GO, KEGG, WikiPathways

RESULTS

ENRICHMENT ANALYSIS (Annotation) :- BioMart (tool)

ensembl_id	symbol	name	entrezgene
ENSG00000165716	DIPK1B	divergent protein kinase domain 1B	138311
ENSG00000165795	NDRG2	NDRG family member 2	57447
ENSG00000165802	NSMF	NMDA receptor synaptonuclear signaling and neuronal migration factor	26012
ENSG00000165914	TTC7B	tetratricopeptide repeat domain 7B	145567
ENSG00000166086	JAM3	junctional adhesion molecule 3	83700
ENSG00000166091	CMTM5	CKLF like MARVEL transmembrane domain containing 5	116173
ENSG00000166681	BEX3	brain expressed X-linked 3	27018
ENSG00000166963	MAP1A	microtubule associated protein 1A	4130
ENSG00000167460	TPM4	tropomyosin 4	7171
ENSG00000167468	GPX4	glutathione peroxidase 4	2879
ENSG00000167641	PPP1R14A	protein phosphatase 1 regulatory inhibitor subunit 14A	94274
ENSG00000168497	CAVIN2	caveolae associated protein 2	8436
ENSG00000168994	PXDC1	PX domain containing 1	221749
ENSG00000169129	AFAP1L2	actin filament associated protein 1 like 2	84632
ENSG00000169247	SH3TC2	SH3 domain and tetratricopeptide repeats 2	79628

Green Annotated genes

RESULTS

ensembl_id	symbol	name	entrezgene
ENSG00000211951	IGHV2-26	immunoglobulin heavy variable 2-26	28455
ENSG00000211952	IGHV4-28	immunoglobulin heavy variable 4-28	28400
ENSG00000211955	IGHV3-33	immunoglobulin heavy variable 3-33	28434
ENSG00000211956	IGHV4-34	immunoglobulin heavy variable 4-34	28395
ENSG00000211959	IGHV4-39	immunoglobulin heavy variable 4-39	28394
ENSG00000211962	IGHV1-46	immunoglobulin heavy variable 1-46	28465
ENSG00000211964	IGHV3-48	immunoglobulin heavy variable 3-48	28424
ENSG00000211965	IGHV3-49	immunoglobulin heavy variable 3-49	28423
ENSG00000211966	IGHV5-51	immunoglobulin heavy variable 5-51	28388
ENSG00000211967	IGHV3-53	immunoglobulin heavy variable 3-53	28420
ENSG00000211970	IGHV4-61	immunoglobulin heavy variable 4-61	28391
ENSG00000211972	IGHV3-66	immunoglobulin heavy variable 3-66	28412
ENSG00000211973	IGHV1-69	immunoglobulin heavy variable 1-69	28461
ENSG00000211976	IGHV3-73	immunoglobulin heavy variable 3-73	28409
ENSG00000213079	SCAF8	SR-related CTD associated factor 8	22828
ENSG00000213585	VDAC1	voltage dependent anion channel 1	7416

Turquoise Annotated genes

RESULTS

ENRICHMENT ANALYSIS (unmapped genes) :-

unmapped_ensembl_id
ENSG00000279673
ENSG00000283633

Green unmapped genes

unmapped_ensembl_id
ENSG00000215067
ENSG00000263264
ENSG00000270157
ENSG00000272316

Turquoise unmapped genes

Why are these genes unmapped?

- ENSEMBEL IDs may correspond to non-coding RNAs or pseudogenes (lack functional annotations or ENTREZ ID mapping in database)
- Outdated ENSEMBEL ID
- Few ENSEMBEL ID maps to multiple ENTREZ IDs or has ambiguous annotations

RESULTS

ENRICHMENT ANALYSIS (green module) :-

Gene_set	Term	Odds Ratio	Combined Score	Genes	P-value	/
GO_Biological_Proce	Nucleosome Assembly (GO:0006334)	18.23405038	202.3289919	H2BC9;H2BC1	1.52E-05	
GO_Biological_Proce	Positive Regulation Of Platelet Activation (GO:0010572)	87.95575221	973.6966866	SELP;GP9;GP9	1.56E-05	
GO_Biological_Proce	Megakaryocyte Development (GO:0035855)	75.38685209	807.9860091	GP9;MPIG6B;	2.21E-05	
GO_Biological_Proce	protein-DNA Complex Assembly (GO:0065004)	16.54154154	176.3318118	H2BC9;H2BC7	2.35E-05	
GO_Biological_Proce	Nucleosome Organization (GO:0034728)	12.75032175	121.0776021	H2BC9;H2BC9	7.51E-05	
GO_Biological_Proce	Positive Regulation Of Cell Activation (GO:0050867)	43.96460177	413.5413538	SELP;GP9;GP9	8.22E-05	
GO_Biological_Proce	Megakaryocyte Differentiation (GO:0030219)	37.68015171	339.6075859	GP9;MPIG6B;	0.000122	
GO_Biological_Proce	Regulation Of Megakaryocyte Differentiation (GO:0045652)	32.96681416	285.7297461	H4C9;FAXDC2	0.000172	
GO_Biological_Proce	Innate Immune Response In Mucosa (GO:0002227)	29.30088496	244.9255349	H2BC12;H2B(	0.000234	
GO_Biological_Proce	Cellular Response To Magnesium Ion (GO:0071286)	116.2631579	932.1127225	SLFN14;RYR3;	0.00033	
GO_Biological_Proce	Myeloid Cell Development (GO:0061515)	21.96902655	166.8814094	GP9;MPIG6B;	0.000502	
GO_Biological_Proce	Regulation Of Platelet Activation (GO:0010543)	21.08920354	157.8970942	SELP;GP9;GP9	0.00056	
GO_Biological_Proce	Mucosal Immune Response (GO:0002385)	21.08920354	157.8970942	H2BC12;H2B(	0.00056	

Gene_set: database (GO_Biological_Process_2023)

Term: Processes or pathway (e.g., "DNA repair")

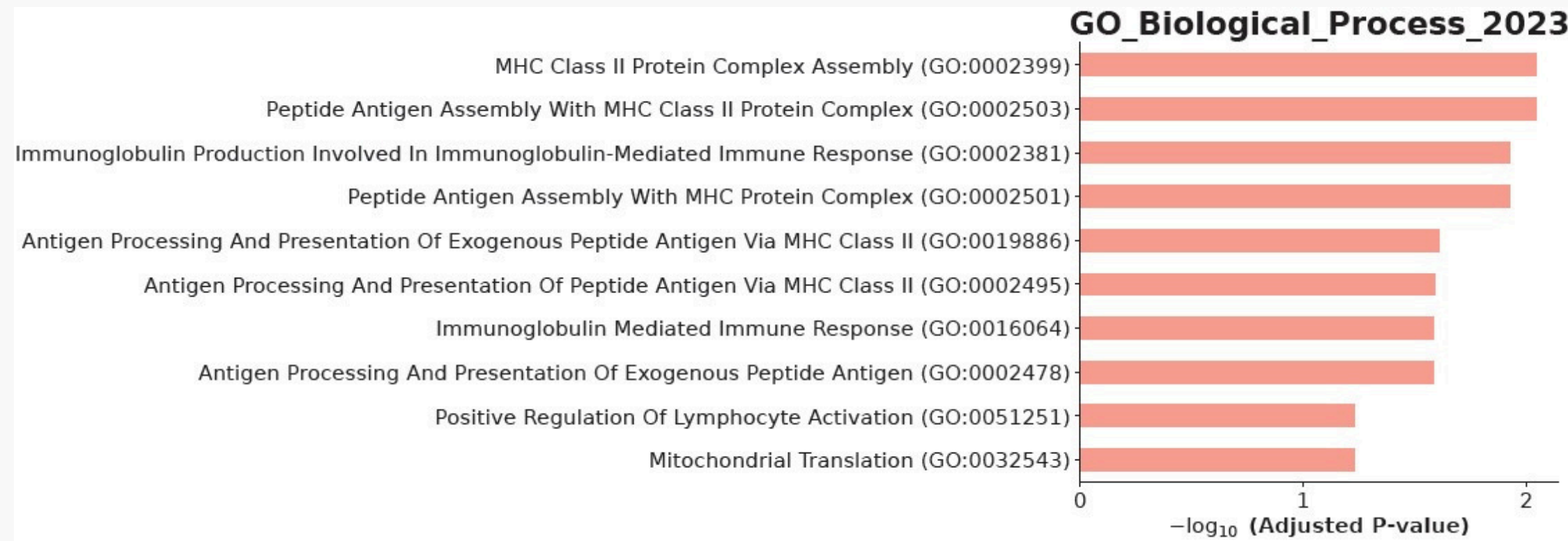**P-value:** The chance the link happened randomly (lower is better).

Odds Ratio: How much more likely your genes are in the process (higher is better).

Combined Score: A ranking combining p-value and odds ratio (higher is better).

Genes: The specific genes from your list contributing to the process.

RESULTS

ENRICHMENT ANALYSIS :-

DotPlot of Gene Ontology (GO) Biological Process (BP) enrichment analysis (green)

- Genes in green module mapped to **ENTREZ IDs** and clusterProfiler identified GO BP terms
- **enriched terms** - genes are functionally realted to immune system processes, **MHC** (**Major Histocompatibility Complex**) **Class II proteins** and **immunoglobulin-mediated** responses.
- Presence of "**Mitochondrial Translation**" suggests some genes play role in **cellular metabolism** or **protein synthesis** within **mitochondria**.
- Reflects biological significance in immune response

RESULTS

GO_Biological_Process_2023
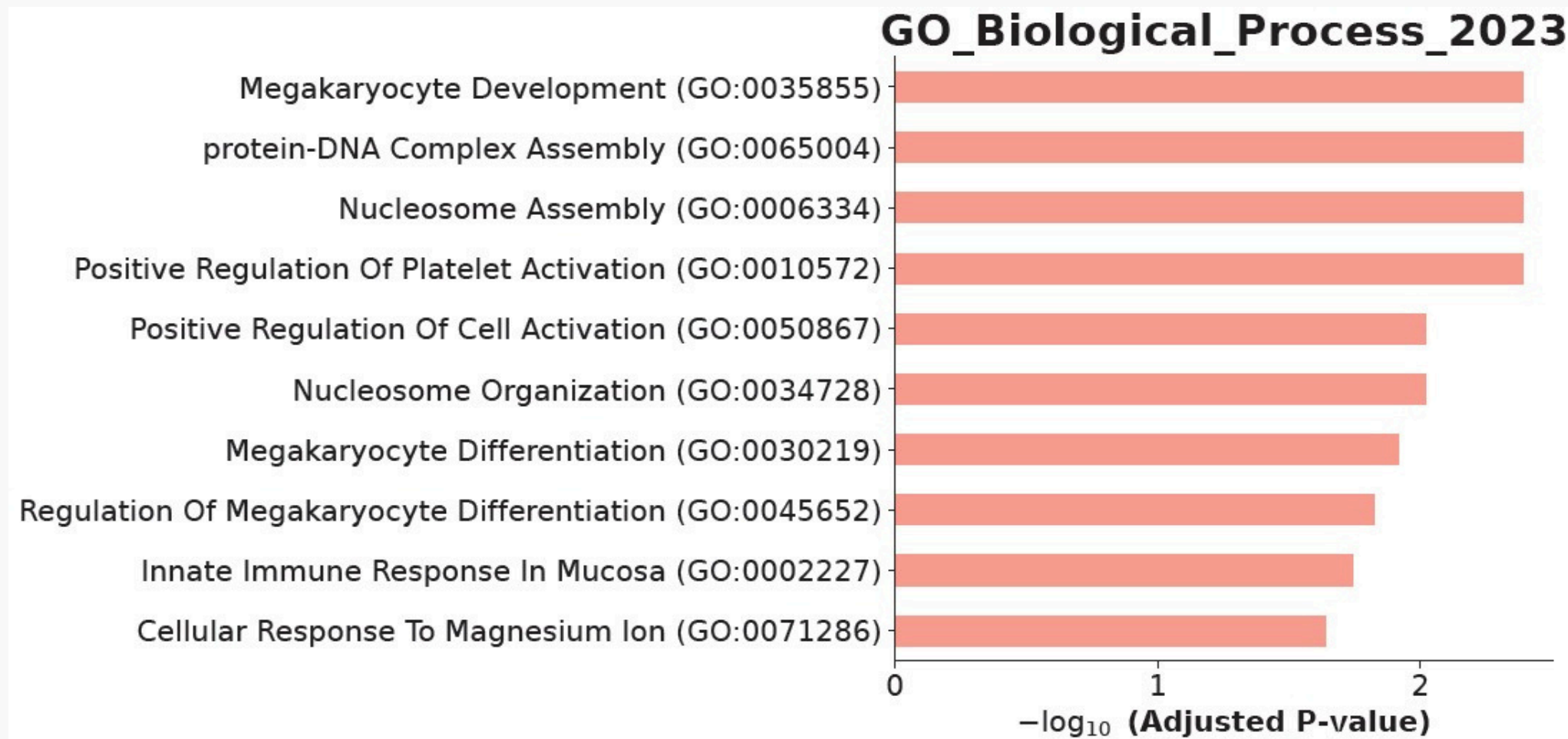
DotPlot of Gene Ontology (GO) Biological Process (BP) enrichment analysis (Turquoise)

- each dot represents an **enriched GO term**, x axis showing significance (**logarithmic p-value**)
- enriched terms highly associated with **megakaryocyte development** and **differentiation** - indicate a biological context related to **hematopoiesis** (**blood cell formation**), **platelet function** or **mucosal immunity**
- This reflects on blood-related disorders (e.g., thrombocytopenia)

RESULTS
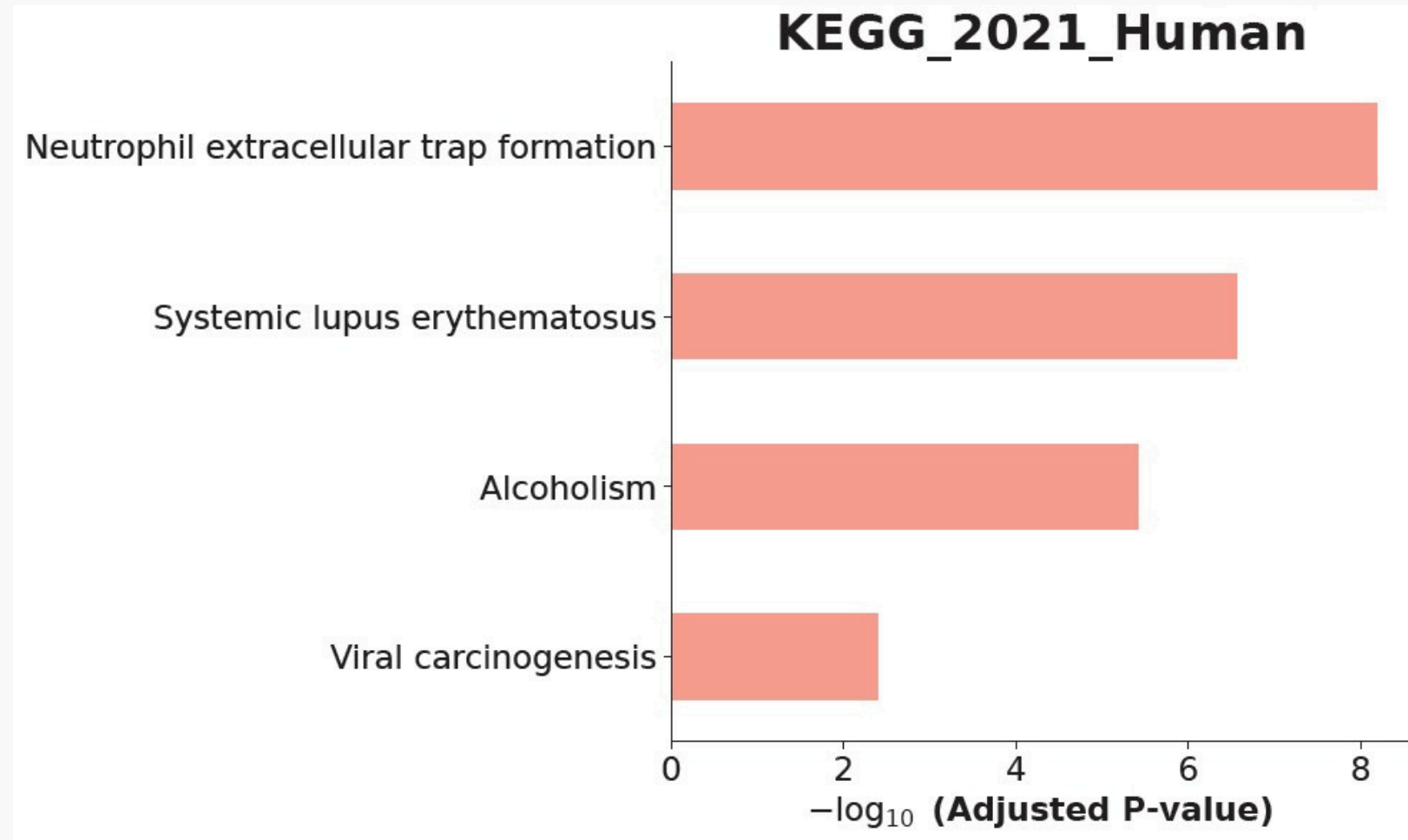
DotPlot of KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway enrichment for human genome (KEGG 2021) (Green)

log10(adjusted P-value) is done by **Hochberg correction** (statistical method used to control **false discovery rate (FDR)**)
- enriched pathways suggest that gene set is associated with **immune-related processes**, **alcohol metabolism** etc.,
- **Neutrophil extracellular trap formation** (highest adjusted P-value) - process where neutrophils release DNA and proteins to trap and kill pathogens
- **Systemic Lupus Erythematosus** - pathway related to systemic lupus erythematosus (**SLE**) involving **immune complex deposition** and **inflammation**

RESULTS

Reactome_2022

DotPlot for Reactome pathway enrichment (2022)

Reactome - manually curated, open-source database of biological pathways and reactions (metabolisms, signaling and gene expression)

Gene set heavily associated with **DNA damage response**, **telomere maintenance**, **chromatin organization**, **transcription regulation**, and **cellular senescence**.

- **Condensation** - during prophase
- Dominance of **DNA repair** suggest gene module is enriched for genes involved in **genomic stability** and **epigenetic control**.

RESULTS
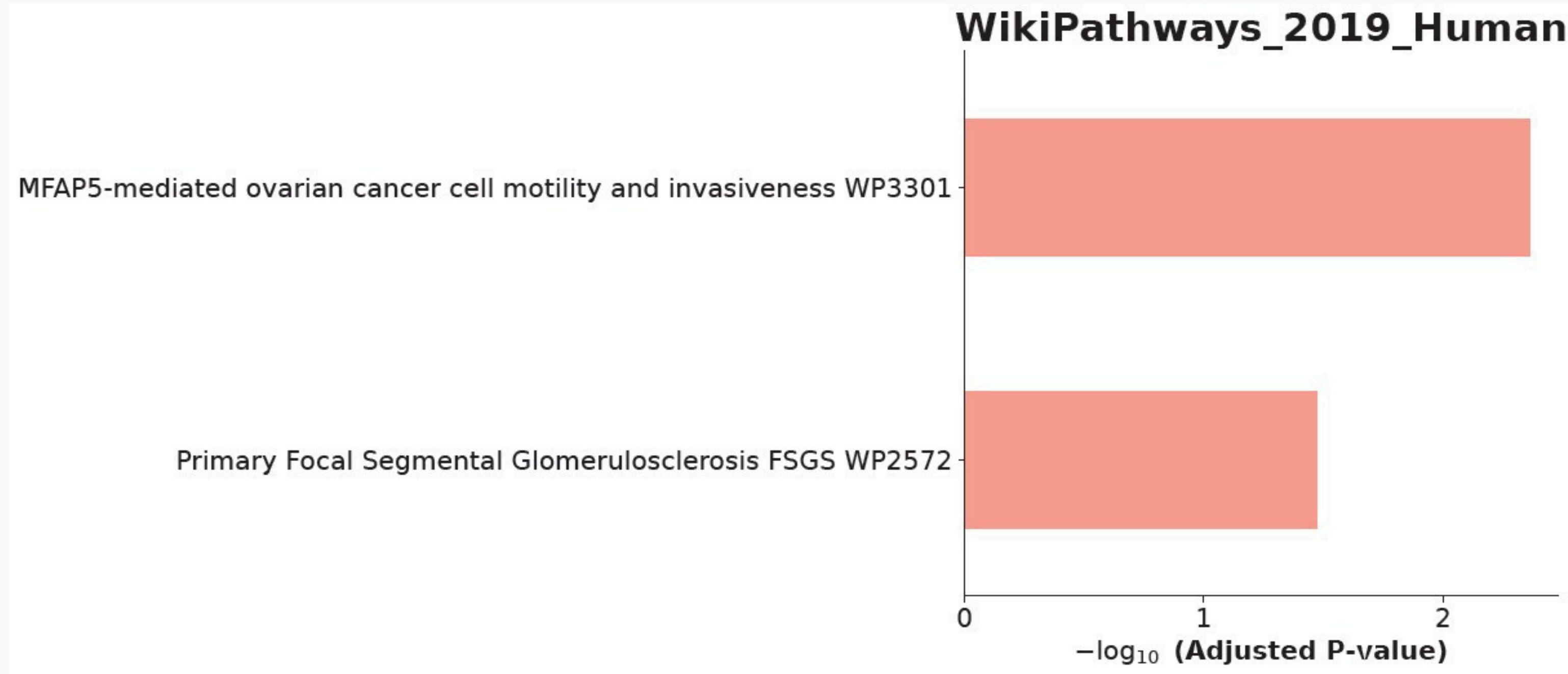
DotPlot for WikiPathways enrichment analysis (2019)

WikiPathways - community curated, open-source database of biological pathways, covering disease, metabolism, and signaling,

INFERENCE

Pathway	Source	Purpose	Useful insights from enrichment analysis
Nucleosome assembly	GO	It organizes DNA into nucleosomes by assembling histone proteins, enabling DNA packaging for cell function	Gene (e.g., **H2BC9** - from histone focus) are key supporting for healthy cell division and gene regulation.
Platelet Activation	KEGG	Triggers platelet aggregation and clotting to stop bleeding and support blood vessel repair	Genes (e.g., **ITGB3** - clotting focus) drive blood clotting, critical for wound healing and linked to disease. Connects to WikiPathways adhesion (ITGB3).
Condensation of prophase chromosomes	WikiPathways	Enables Ovarian cancer cells to move and invade tissues via adhesion and signaling	Genes(**ITGB3** - promote cancer spread). ITGB3 link to KEGG's cancer, Reactome's motility, GO's adhesion

CONCLUSION

The project began with **constructing a co-expression gene network** and identified outliers and hub genes.

The project further continued with **enrichment analysis from GO, KEGG, Reactome and WikiPathways** to reveal gene list plays a crucial role in DNA management, blood clotting and disease processes.

Collectively these findings suggest that **genes are key players in DNA organization**, **hemostasis**, and **disease pathways**, offering insights for research.

RESEARCH PAPERS RELATED TO ETHICS

ETHICS

Informed Consent in Genomics and Genetic Research (2021)
-National Heath Institute (NIH) https://pmc.ncbi.nlm.nih.gov/articles/PMC3216676/

Ethical Considerations in Research with Genomic Data
-The New Bioethics, a multidisciplinary Journal of Biotechnology and the Body
https://doi.org/10.1080/20502877.2022.2060590

Challenges and potential solutions to health disparities in genomic medicine
-ScienceDirect; https://doi.org/10.1016/j.cell.2022.05.010

Ethical Challenges Posed by Big Data
National Center for Biotechnology Information (NCBI), National library of Medicine
https://pmc.ncbi.nlm.nih.gov/articles/PMC7819582/

PATENTS RELATED TO ETHICS

ETHICS

Method for Diagnosing Lung Cancers Using Gene Expression Profiles in Peripheral Blood Mononuclear Cells
Patent No: US 20200191791A1; https://patents.google.com/patent/US8476420B2/en

Method for excavating colorectal cancer radiotherapy specific genes by using weight gene co-expression network
Patent No: CN 109872772B; https://patents.google.com/patent/CN109872772B/en?q=(weighted+gene+expression+network+analysis)&language=ENGLISH

Method for Diagnosing Colorectal Cancer from a Human Blood Sample
Patent No: US 20170096709A1;
https://patentimages.storage.googleapis.com/ef/72/47/71937553569f51/US20170096709A1.pdf

Gene expression profiling for the diagnosis of prostate cancer
Patent No: US20160340745A1; https://patents.google.com/patent/US20160340745A1/en

Thank you