Differential Gene Expression Analysis Between Caucasian and African Hair Types

An In-depth Comparative Study Using GEO2R, g:Profiler, STRING, and Cytoscape

Author: *Smile Vij*

Date: *August 8, 2024*

Introduction

This study investigates differential gene expression in hair follicles between individuals with very curly and straight hair, utilizing high-throughput microarray analysis to identify key genetic differences that influence hair shape.

Objectives:

- To identify differentially expressed genes between Caucasian and African hair types.

- To analyze and interpret the gene expression data using various bioinformatics tools.

- To create a network of all the top differentially expressed genes.
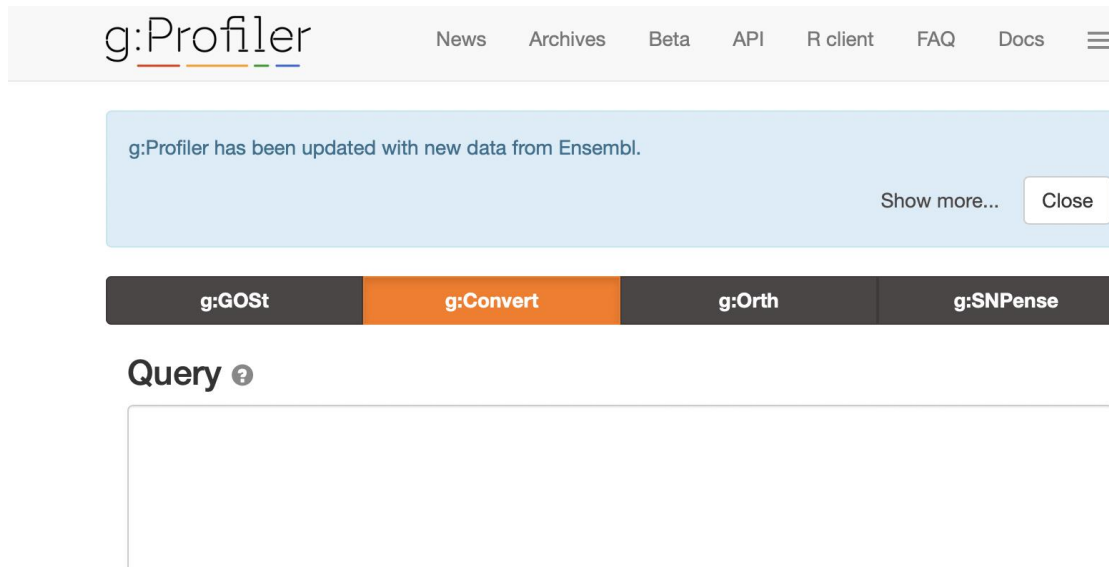
Titles and Experiment Information

- Status: Public on Feb 19, 2024
- Title: Differences in gene expression between very curly and straight human hair
- Organism: Homo sapiens
- Experiment type: Expression profiling by array

- Summary: Hair shape is defined in the follicle: large hair follicles produce 'terminal' hairs, small follicles produce fine 'vellus' hairs, curved follicles produce curly hair in all ethnicities, according to the available literature. In order to understand better how shape is determined in the follicle and complement the publicly available knowledge on hair shape genetics, our research group undertook a global high-throughput approach to compare the levels of gene expression among straight and very curly hair. The microarray was the methodology chosen. A list of the transcripts differently expressed that pass the statistical two independent groups test including correction for multiple testing (Bonferroni-Hochberg with 0.05 cut-off) was obtained. From 85 genes that were significantly different 68 genes were more expressed in very curly hair follicles and 17 genes were more expressed in straight hair follicles.

- Overall design: The team collected 50 hair follicles from each volunteer: 6 donors with extremely curly hair and 5 donors with straight hair, all non-bald adult males with ages between 19-40 years old. The isolated RNA was processed and used on Affymetrix GeneChip HuGene 1.0 ST Arrays (Affymetrix Core Facility at Instituto Gulbenkian de Ciencia, Oeiras, Portugal)

- Web link: https://www.sciencedirect.com/science/article/pii/S2090123223003508?via%3Dihub

GEO2R Analysis

- Objective: To perform differential expression analysis using GEO2R.

- Sample Groups: 6 Caucasian hair samples and 5 African hair samples.

- Analysis: Each gene pair was analyzed separately, e.g., 12-12, 14-14, etc.

- Link:https://www.ncbi.nlm.nih.gov/geo/geo2r/?acc=GSE19393 83.

- Description: GEO2R is an interactive web tool that allows users to compare two or more groups of samples in a GEO series to identify differentially expressed genes.

g:Profiler Analysis

- Objective: To convert microarray data identifiers to those usable in STRING analysis.

- Link: https://biit.cs.ut.ee/gprofiler/gost

- Description: g:Profiler is a web server for functional enrichment analysis and gene identifier conversion.

Log Fold Change Analysis

- Objective: To determine the regulation patterns of genes.

- Method: Sorting genes based on log fold change (logFC) values to identify upregulated and downregulated genes.

- Results: Top 100 regulated genes for both upregulated and downregulated categories were identified.

Compilation of Data

Process:

- Compiled all the identified genes from 10 analyses into a single Excel sheet.

- Removed non-identifiable genes to retain relevant data.

Gene Functions and Literature Review

- With the Up regulated and Down regulated gene types being analysed and found out and the next step is to proceed with the literature study. Pertaining to the analysis, the most significant types in the up and down regulated genes are denoted and are displayed in the next slide with their functions.

UP REGULATED GENES

ADAT1
Involved in tRNA modification.
Link:
https://www.genecards.org/cgi-bin/carddisp.pl?gene=ADAT1

BRAF
Encodes a protein involved in sending signals inside cells, directing cell growth.
Link: https://www.genecards.org/cgi-bin/carddisp.pl?gene=BRAF

CMKLR2-AS
Antisense RNA associated with the regulation of CMKLR2.
Link: https://www.genecards.org/cgi-bin/carddisp.pl?gene=CMKLR2-AS

DSC3
Involved in cell adhesion.
Link:
https://www.genecards.org/cgi-bin/carddisp.pl?gene=DSC3

SCUBE3
Plays a role in cell signaling.
Link:
https://www.genecards.org/cgi-bin/carddisp.pl?gene=SCUBE3

SLC4A9
Involved in bicarbonate transport.
Link:
https://www.genecards.org/cgi-bin/carddisp.pl?gene=SLC4A9

SERP1
Involved in stress response and protein processing.
Link:
https://www.genecards.org/cgi-bin/carddisp.pl?gene=SERP1

SLC25A35
Mitochondrial transporter protein.
Link:
https://www.genecards.org/cgi-bin/carddisp.pl?gene=SLC25A35

TMEM40
Transmembrane protein with unknown function.
Link:
https://www.genecards.org/cgi-bin/carddisp.pl?gene=TMEM40

PADI1
Involved in protein deimination.
Link:
https://www.genecards.org/cgi-bin/carddisp.pl?gene=PADI1

TNFAIP8
Involved in apoptosis regulation.
Link:
https://www.genecards.org/cgi-bin/carddisp.pl?gene=TNFAIP8

MMD
Plays a role in the immune response.
Link:
https://www.genecards.org/cgi-bin/carddisp.pl?gene=MMD

TRIB3
Regulates cell survival and stress responses.
Link:
https://www.genecards.org/cgi-bin/carddisp.pl?gene=TRIB3

TVP23A
Involved in vesicle trafficking.
Link:
https://www.genecards.org/cgi-bin/carddisp.pl?gene=TVP23A

GLUD2
Encodes mitochondrial glutamate dehydrogenase.
Link:
https://www.genecards.org/cgi-bin/carddisp.pl?gene=GLUD2

FGD6
Involved in actin cytoskeleton organization.
Link:
https://www.genecards.org/cgi-bin/carddisp.pl?gene=FGD6

FAM247B
Family with sequence similarity 247, member B.
Link:
https://www.genecards.org/cgi-bin/carddisp.pl?gene=FAM247B

FAM247A
Family with sequence similarity 247, member A.
Link:
https://www.genecards.org/cgi-bin/carddisp.pl?gene=FAM247A

ELMO2
Plays a role in cell migration and cytoskeletal organization.
Link:
https://www.genecards.org/cgi-bin/carddisp.pl?gene=ELMO2

DOWN REGULATED GENES

ODC1 — Involved in polyamine biosynthesis, essential for cell growth.
Link: https://www.genecards.org/cgi-bin/carddisp.pl?gene=ODC1

AP1S2 — Associated with X-linked intellectual disability.
Link: https://www.genecards.org/cgi-bin/carddisp.pl?gene=AP1S2

C5orf60 — Function is currently unknown.
Link: https://www.genecards.org/cgi-bin/carddisp.pl?gene=C5orf60

CAPRIN1 — Involved in RNA binding and associated with stress granules.
Link: https://www.genecards.org/cgi-bin/carddisp.pl?gene=CAPRIN1

CCDC136 — Involved in the cellular process, but specific function is unknown.
Link: https://www.genecards.org/cgi-bin/carddisp.pl?gene=CCDC136

CEP126 — Plays a role in microtubule organization.
Link: https://www.genecards.org/cgi-bin/carddisp.pl?gene=CEP126

CTNNA2 — Encodes alpha-catenin, involved in cell adhesion.
Link: https://www.genecards.org/cgi-bin/carddisp.pl?gene=CTNNA2

EYA4 — Involved in eye and ear development.
Link: https://www.genecards.org/cgi-bin/carddisp.pl?gene=EYA4

FOXCUT — Long non-coding RNA associated regulation of gene expression.
Link: https://www.genecards.org/cgi-bin/carddisp.pl?gene=FOXCUT

HNRNPD — Involved in mRNA stability and translation regulation.
Link: https://www.genecards.org/cgi-bin/carddisp.pl?gene=HNRNPD

MCM7 — Component of the MCM2-7 complex involved in the initiation replication.
Link: https://www.genecards.org/cgi-bin/carddisp.pl?gene=MCM7

USP25 — Encodes a deubiquitinating enzyme.
Link: https://www.genecards.org/cgi-bin/carddisp.pl?gene=USP25

UBE2D1 — Involved in ubiquitin-proteasome system.
Link: https://www.genecards.org/cgi-bin/carddisp.pl?gene=UBE2D1

SIDT2 — Plays a role in RNA transport.
Link: https://www.genecards.org/cgi-bin/carddisp.pl?gene=SIDT2

RUBCN — Involved in autophagy regulation.
Link: https://www.genecards.org/cgi-bin/carddisp.pl?gene=RUBCN

OGT — Involved in protein glycosylation.
Link: https://www.genecards.org/cgi-bin/carddisp.pl?gene=OGT

NOX5 — Produces reactive oxygen species.
Link: https://www.genecards.org/cgi-bin/carddisp.pl?gene=NOX5

MST1R — Encodes a receptor tyrosine kinase involved in cell growth.
Link: https://www.genecards.org/cgi-bin/carddisp.pl?gene=MST1R

MIR93 — MicroRNA involved in the regulation of gene expression.
Link: https://www.genecards.org/cgi-bin/carddisp.pl?gene=MIR93

MIR25 — MicroRNA involved in the regulation of gene expression.
Link: https://www.genecards.org/cgi-bin/carddisp.pl?gene=MIR25

MIR106B — MicroRNA involved in the regulation of gene expression.
Link: https://www.genecards.org/cgi-bin/carddisp.pl?gene=MIR106B

STRING and Cytoscape Analysis

- Objective: To visualize protein-protein interactions.
- Tools Used:
- STRING: Setting interaction score to low and selecting 5 interactors.
- Link: https://string-db.org/
- Cytoscape: For detailed network analysis.
- Link: https://cytoscape.org/

Results from STRING

Downregulated Network:
•**Decreased Expression:** Network of proteins with reduced expression.
•**Clustering:** Identifies potential co-regulated pathways and suppressed biological processes.
•**Key Insight:** Highlights pathways less active under the studied conditions, suggesting potential areas of functional suppression.

Upregulated Network:
•**Increased Expression:** Network of proteins with elevated expression.
•**Clustering:** Reveals groups of proteins likely involved in activated biological processes.
•**Key Insight:** Identifies critical pathways and processes that are upregulated, pointing to possible therapeutic targets.

Visuals: Network diagrams showing interactions between the differentially expressed genes.

Downregulated network of string with clustering:
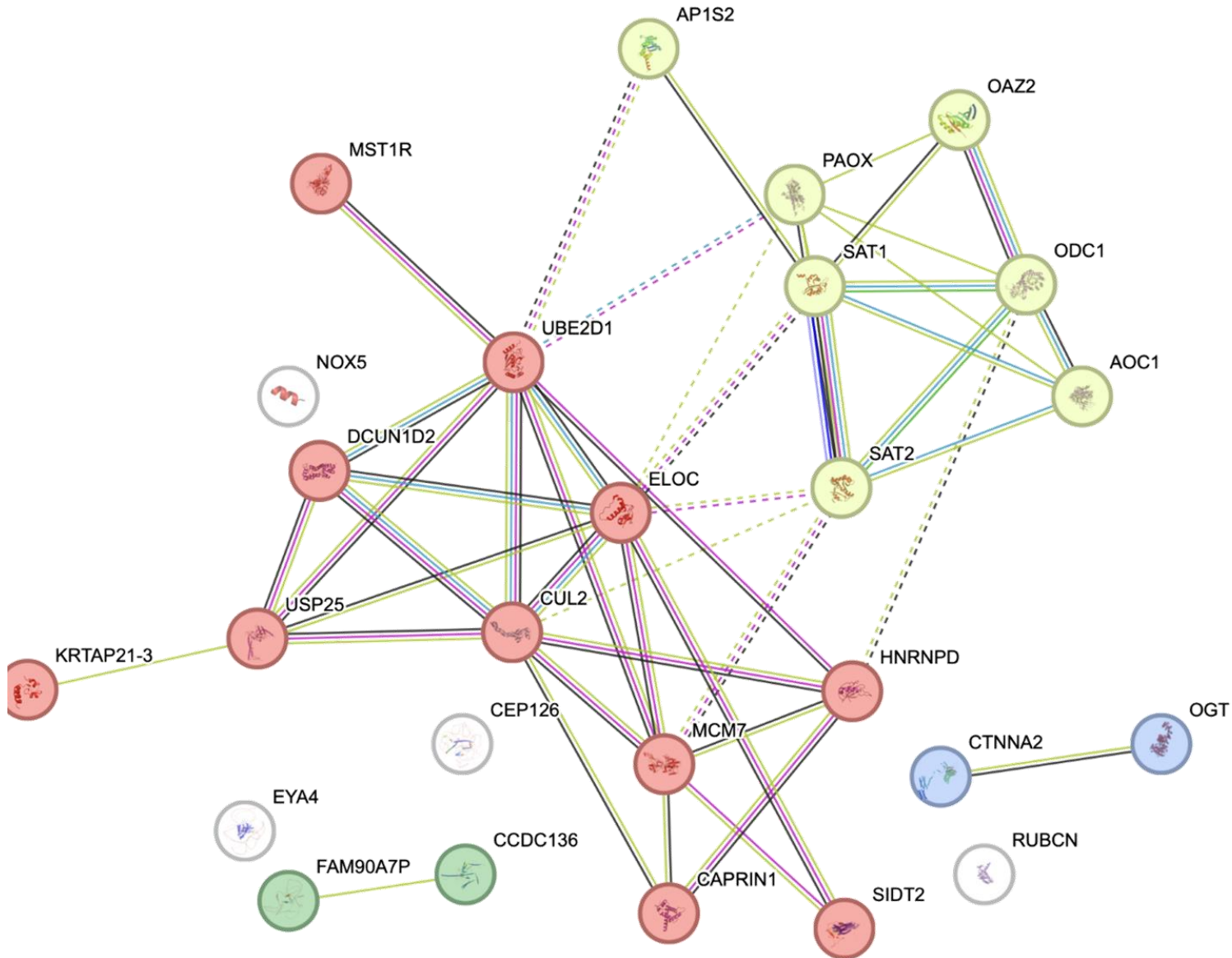
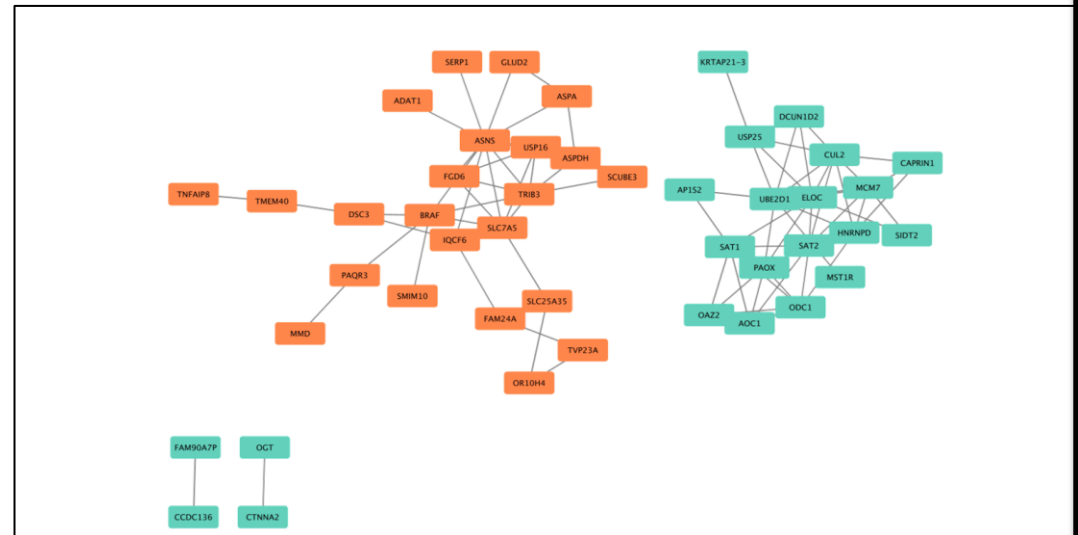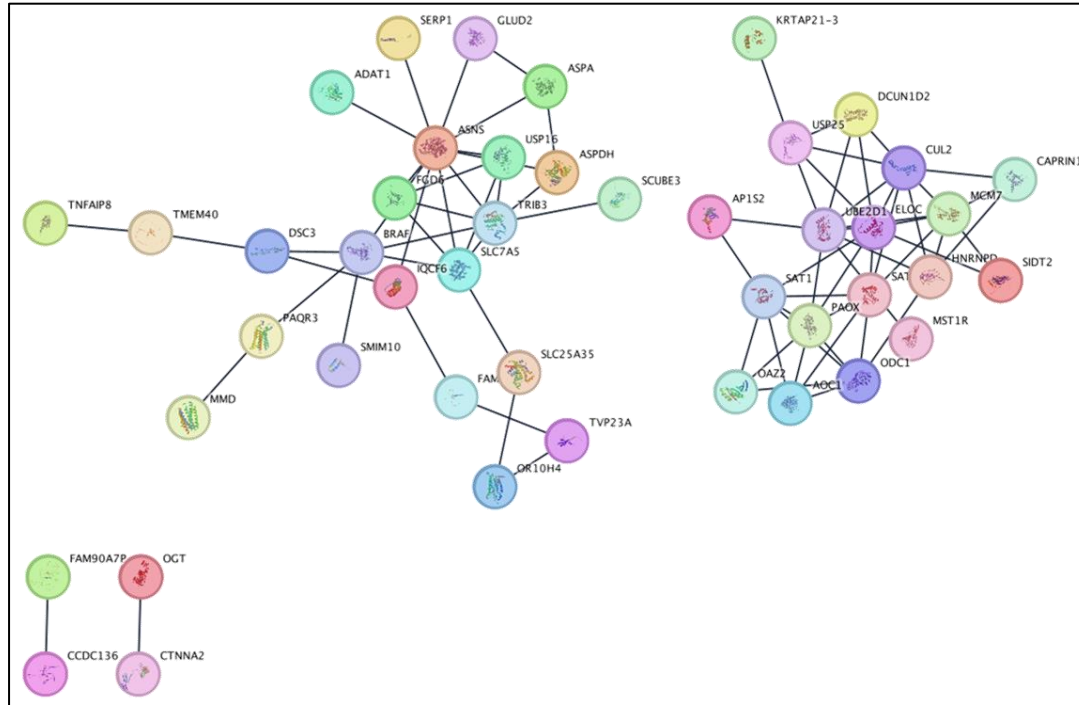# Upregulated network of string with clustering:
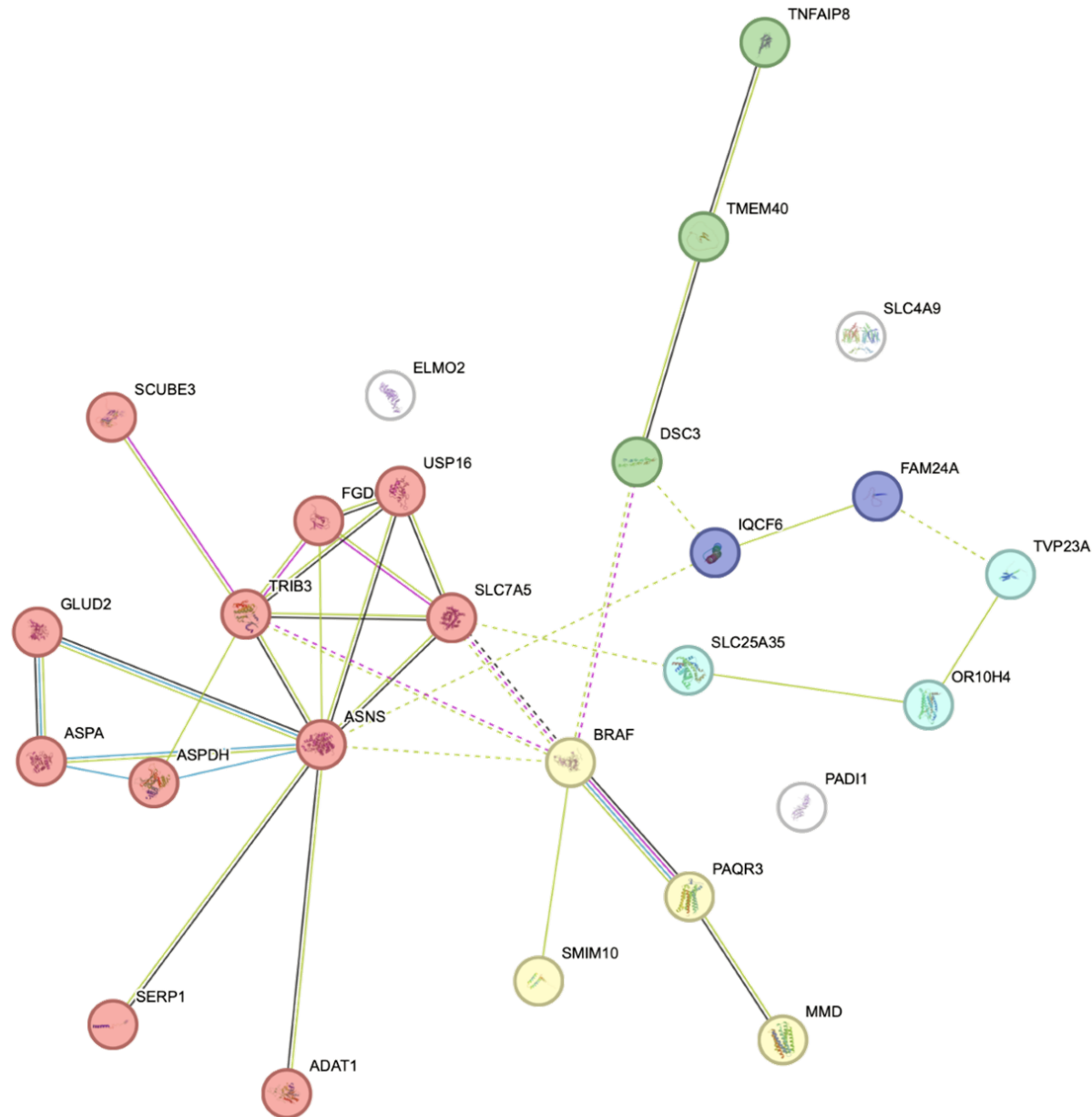
Analysis of Cytoscape with String for Down Regulated gene types
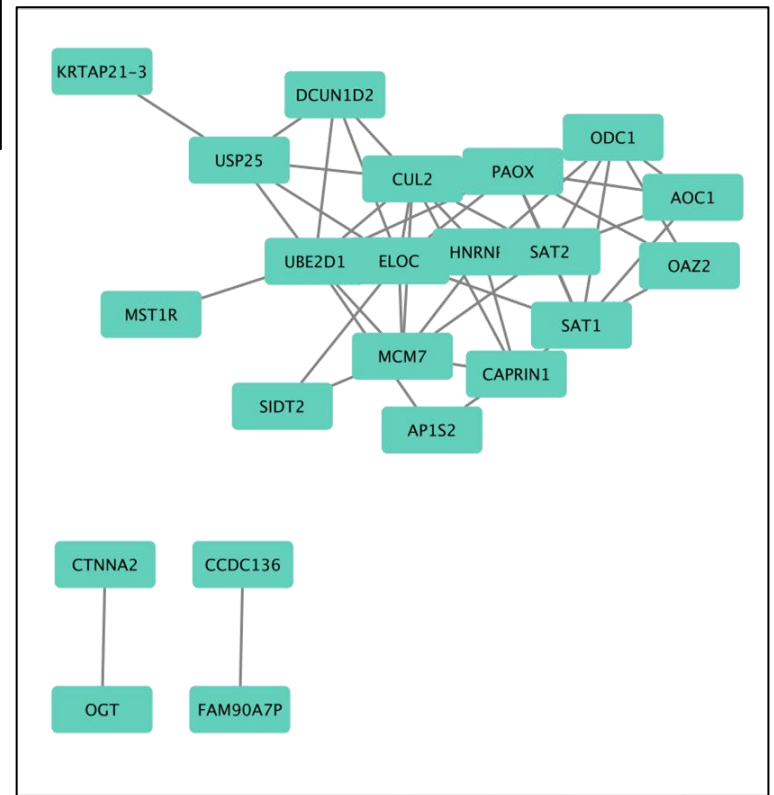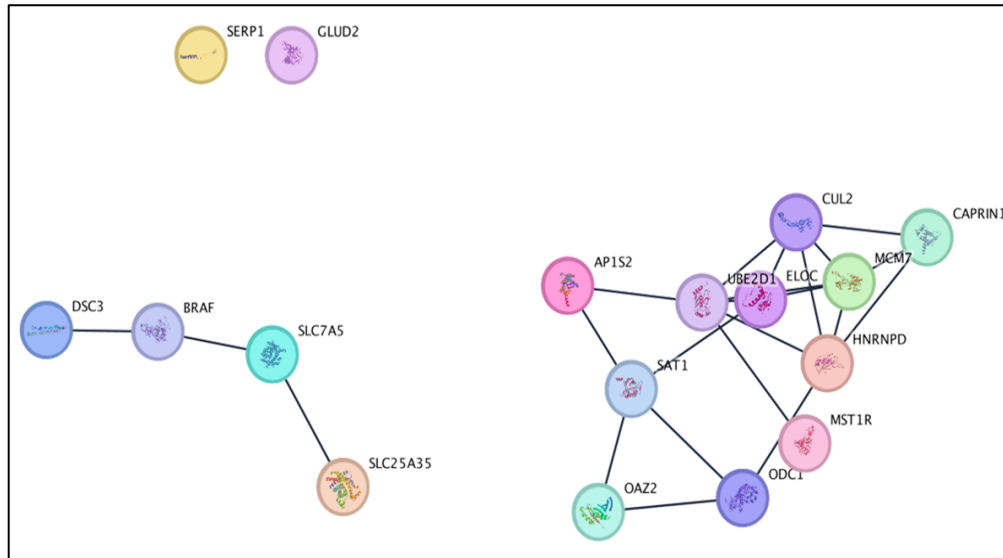
Analysis of Cytoscape with String for Up Regulated gene types
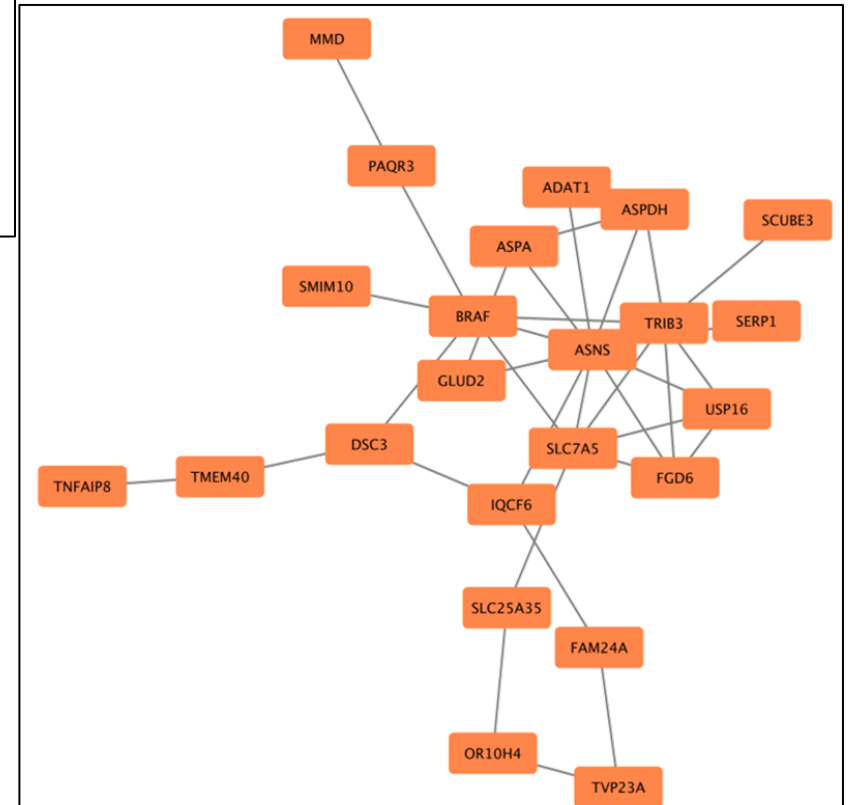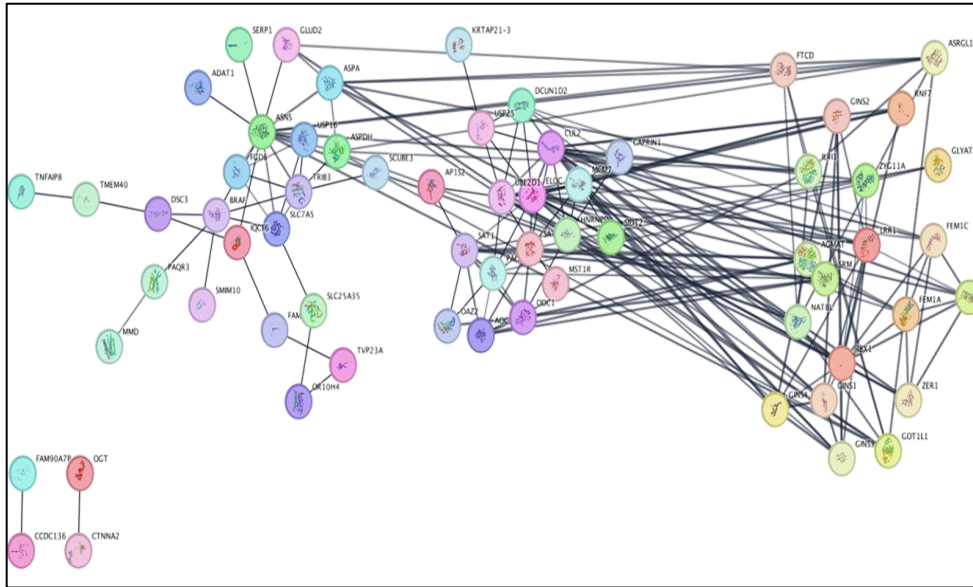
Analysis of Cytoscape with String for Merged gene types

Conclusions

Findings:

- Significant differences in gene expression between Caucasian and African hair types.

- Identification of key genes involved in hair type differentiation.

Future Directions:

- Further validation of identified genes through experimental studies.

- Exploration of potential therapeutic targets for hair-related conditions.

References

References:

- GEO2R, g:Profiler, STRING, Cytoscape, and relevant literature.
- https://www.sciencedirect.com/science/article/pii/S2090123223003508?via%3Dihub
- https://www.ncbi.nlm.nih.gov/geo/geo2r/?acc=GSE193983.
- https://biit.cs.ut.ee/gprofiler/gost
- https://string-db.org/
- https://cytoscape.org/
- https://string-db.org/cgi/network?taskId=bUtIb4VcTSA1&sessionId=bDDkNAEgDT8J
- https://string-db.org/cgi/network?taskId=bk1bdYWVlz6I&sessionId=b2OzD2YrFcxd
- https://www.genecards.org/cgi-bin/carddisp.pl?gene=ADAT1