DEG Induced by Betel Nut Chewing: RNA Seq Analysis Link to Type 2 Diabetes and Obesity

Abstract

Betel nut chewing is the fourth most common addictive habit in the world. People who are addicted with this habit are prone to Type 2 Diabetes(T2D) and Obesity. This project is mainly focused on finding out the possible genes that are included during the for T2D and Obesity through a transcriptomics approach. The datasets used for this project were taken from an existing project in the GEO database, where human monocyte cell lines were incubated in Arecoline and MNPA, which are products from Betel nut thought to be associated with T2D and Obesity. I have attempted to recreate the transcriptomics findings using my own analysis pipeline and generated output. At the conclusion of the analysis, genes and their associated metabolic pathways were identified that have the potential to contribute to the development of type 2 diabetes and obesity. The analysis revealed that various metabolic pathways were found to be altered due to the genes that were identified as having potential to cause these conditions.

Acknowledgements

I would like to express my deepest gratitude to my advisor, Ms. Shilpa Rathode, for their unwavering support, guidance, and encouragement throughout the course of this research. Their insightful feedback and expertise were invaluable in shaping this thesis. I am immensely grateful to the Mediomix Diagnosis and Bioresearch Centre for providing me with the opportunity to conduct my research in their facility. The employees at Mediomix taught me the entire process and supported me throughout the completion of this project. Their expertise and assistance were instrumental in the successful execution of this research. Special thanks to my lab colleagues and friends, whose camaraderie and collaboration made this journey enjoyable and enriching. I am grateful to my family for their constant love, patience, and encouragement, which motivated me to persevere through challenges.

I extend my gratitude to the NCBI-GEO database for providing the datasets used in this study. Reproducing the RNA-seq analysis from the research paper "Areca catechu-(Betel-nut)-induced whole transcriptome changes in a human monocyte cell line that may have relevance to diabetes and obesity; a pilot study" by Shirleny R. Cardosa et al. was a critical component of my research, and I appreciate the accessibility and quality of the data provided.

Introduction

Type 2 diabetes has rapidly emerged as a significant public health challenge in South Asia in recent decades. During this period, major lifestyle transformations associated with economic development, industrialization, urbanization, and globalization have been key drivers in the increasing burden of non-communicable diseases in the region. A decline in dietary quality, reduced physical activity, and increased sedentary behaviors have contributed to the rising prevalence of type 2 diabetes and related risk factors in the South Asian population. According to the International Diabetes Federation's 2017 estimates, the prevalence of diabetes in adults in South Asian countries ranges from 4.0% in Nepal to 8.8% in India. The prevalence of overweight varies from 16.7% in Nepal to 26.1% in Sri Lanka, while the prevalence of obesity ranges from 2.9% in Nepal to 6.8% in Sri Lanka. An increasing proportion of children, adolescents, and women in the region are overweight or obese, leading to a heightened risk of type 2 diabetes. Ethnic South Asians, including Indians, exhibit greater metabolic risk at lower levels of BMI compared to other ethnic groups, with type 2 diabetes often developing at a younger age and progressing more rapidly in terms of diabetic complications. Given the presence of multiple risk factors and a body composition predisposing to type 2 diabetes, South Asians, including the Indian population, should be a key target for aggressive prevention efforts.

A synthesis of 17 studies conducted in Asian populations revealed a link between chewing betel quid (BQ) and heightened risks of obesity, diabetes, metabolic syndrome, cardiovascular disease, and overall mortality. These results hold significance for metabolic disease prevention efforts, particularly given the escalating rates of these conditions in South-East Asia and the Western Pacific region.

In a study conducted, it was observed that individuals who chewed areca nut exhibited higher age-adjusted prevalence rates of hyperglycemia (11.4% versus 8.7%) and Type 2 diabetes (10.3% versus 7.8%) compared to non-chewers. Furthermore, areca nut chewing was found to independently elevate the risk of hyperglycemia (adjusted odds ratio [OR] 1.19, 95% CI 0.97–1.45) and Type 2 diabetes (adjusted OR 1.29, 95% CI 1.04–1.60). The study also revealed a dose-dependent relationship between the duration and frequency of chewing and the risk of Type 2 diabetes, as well as hyperglycemia. Specifically, the risk increased with longer durations of chewing (adjusted OR 1.32 for a duration of 10–19 years and 1.41 for a duration of ≥20 years) and higher daily consumption rates of areca nut (adjusted OR 1.14 for <10 pieces/day, 1.30 for 10–19 pieces/day, and 2.02 for ≥20 pieces/day), with similar trends observed for hyperglycemia.
Also, it was found that the age-adjusted prevalence of metabolic syndrome was notably higher among current chewers (25.13%), followed by ex-chewers (22.04%), and lowest

among non chewers (15.73%) ($P < 0.0001$). Adjusted odds ratios (95% confidence intervals) for metabolic syndrome were 1.38 (1.19, 1.60) and 1.78 (1.53, 2.08) in ex-chewers and current chewers, respectively, after accounting for other significant factors like a family history of hypertension and diabetes mellitus. Furthermore, significant odds ratios for individual components of metabolic syndrome ranged from 1.24 for hyperglycemia (95% CI: 1.09, 1.64) to 1.90 (95% CI: 1.66, 2.19) for hypertriacylglycerolemia. The study also revealed a dose-response relationship, indicating increasing odds ratios for metabolic syndrome with higher consumption rates of betel quid, regardless of the rate, duration, or cumulative exposure of use.

Methodology

From the Study

RNA Sequencing and Bioinformatics Analysis

This study utilized RNA-sequencing and bioinformatics approaches to analyze gene expression profiles. RNA libraries were constructed from 100 ng of total RNA using the NEBNext Ultra II kit, incorporating polyA isolation. The quality and quantity of the RNA libraries were assessed using the Qubit dsDNA HS assay and Agilent 2100 Bioanalyzer, respectively. Indexed paired-end libraries were sequenced on an Illumina NextSeq 500 platform at the Queen Mary University of London Genome Centre.

Data Processing and Analysis

Sequenced reads were mapped using Kallisto, followed by differential gene expression analysis with Sleuth, employing a generalized linear model and bootstrap estimation of inferential variance. Genome-wide p-values were corrected using the Bonferroni adjustment method.

Functional Annotation and Pathway Enrichment

Functional annotation and pathway enrichment analyses were carried out using DAVID, Reactome, and Metascape tools. These methodologies facilitated an in-depth exploration of gene expression alterations and the biological pathways associated with the experimental conditions, yielding significant insights into the underlying molecular mechanisms.

My Analysis

Data Collection

The raw sequencing data from all six samples were deposited in the SRA database. Three samples were treated with Arecoline (SRR14995014, SRR14995015, SRR14995016), and the remaining three were treated with MNPA (SRR14995017, SRR14995018, SRR14995019). These datasets are accessible under the GEO accession number GSE179143. The human genome assembly GRCh38.p14 from the Ensembl database served as the reference genome. Additionally, the corresponding gene transcript annotation from Ensembl was utilized for downstream analysis with the Cuffdiff tool.

Data Pre-Processing

Adaptor sequences were removed using Fastp v0.23.2, and the quality of the trimmed reads was assessed using FastQC v0.11.9 and aggregated with MultiQC v1.12.

Alignment

Alignment was performed using Hisat2 v2.2.1. The reference human genome was indexed with the hisat2-build utility. Processed reads were then aligned to this indexed genome using Hisat2 with the --dta-cufflinks option, ensuring compatibility with downstream analysis tools like Cuffdiff.

Downstream Processing

Post-alignment, BAM files were subjected to downstream analysis to identify significant and non-significant genes using Cuffdiff v2.2.1. The resulting gene_exp.diff file contained detailed information on test_id, gene_id, gene, locus, log2(fold_change), test_stat, p_value, q_value, and significance. Data analysis focused on genes with a p_value < 0.005 to determine significance.

Pathway Enrichment

Pathway enrichment analysis was conducted using DAVID from NCBI and ShinyGO 0.8. Pathways with an enrichment score greater than 4 were selected for further interpretation and conclusion.

Results

From the paper

In this study, transcriptomics was employed to investigate the genetic responses triggered by arecoline and MNPA, focusing on their implications for diabetes, obesity, and metabolic syndrome. The analysis identified significant gene hits following incubation with each substance, with subsequent refinement based on log-fold change criteria, yielding 15 genes for arecoline and 39 genes for MNPA. Independent evaluation by researchers using databases like GeneCards and PubMed assessed the relevance of these genes to metabolic disorders. Key findings included potential biomarkers such as IGFB3, CLEC10A, and JUP for arecoline, and GLDN, GRIP1, NEGR1, among others, for MNPA. Pathway analysis revealed enrichment in pathways related to cell adhesion, apoptosis, and immune response regulation, shedding light on the underlying biological mechanisms influenced by these substances. Comparison between arecoline and MNPA effects highlighted H3F3AP4 as the sole overlapping gene under strict criteria, indicating shared molecular pathways potentially linked to disease mechanisms. This transcriptomic approach offers valuable insights into the molecular responses to arecoline and MNPA, suggesting avenues for further research into their roles in metabolic disorders.

In this study, transcriptomics was employed to investigate the genetic responses triggered by arecoline and MNPA, focusing on their implications for diabetes, obesity, and metabolic syndrome. The analysis identified significant gene hits following incubation with each substance, with subsequent refinement based on log-fold change criteria, yielding 15 genes for arecoline and 39 genes for MNPA. Independent evaluation by researchers using databases like GeneCards and PubMed assessed the relevance of these genes to metabolic disorders. Key findings included potential biomarkers such as IGFB3, CLEC10A, and JUP for arecoline, and GLDN, GRIP1, NEGR1, among others, for MNPA. Pathway analysis revealed enrichment in pathways related to cell adhesion, apoptosis, and immune response regulation, shedding light on the underlying biological mechanisms influenced by these substances. Comparison between arecoline and MNPA effects highlighted H3F3AP4 as the sole overlapping gene under strict criteria, indicating shared molecular pathways potentially linked to disease mechanisms. This transcriptomic approach offers valuable insights into the molecular responses to arecoline and MNPA, suggesting avenues for further research into their roles in metabolic disorders.

My Results

Data Quality Assessment: Pre- and Post-Trimming Analysis

Quality control (QC) is vital in RNA-sequencing (RNA-seq) workflows to ensure reliable downstream analyses. Twelve RNA-seq samples underwent QC analysis before and after adapter trimming using FastQC and aggregated with MultiQC. This report summarizes the key findings.

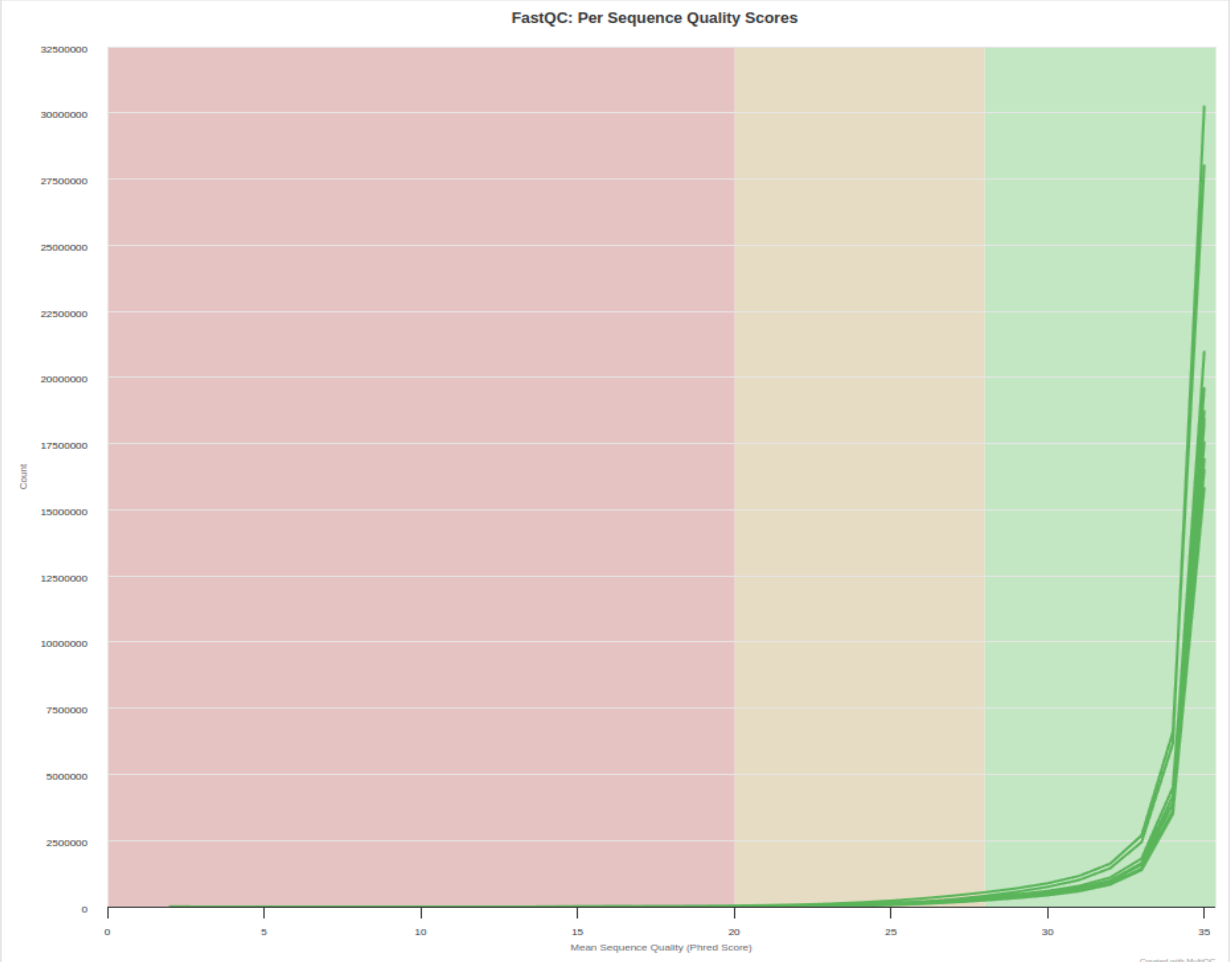
Pre-Trimming Quality Assessment

Per Base Sequence Quality - Before trimming, most bases had Phred scores above 30, indicating high-quality reads. However, quality dropped slightly at the read ends.
Per Sequence GC Content - GC content across samples was consistent and within the expected range for human RNA-seq data (40-60%), suggesting no significant contamination.
Adapter Content - High levels of adapter sequences were present, particularly at the read ends, indicating the need for trimming.
Sequence Duplication Levels - Moderate sequence duplication was observed, which can indicate PCR amplification biases.
Kmer Content - Enrichment of specific sequences, likely from overrepresented adapters or contaminants, was detected.

Post-Trimming Quality Assessment

Per Base Sequence Quality - Post-trimming, Phred scores improved significantly, with nearly all bases scoring above 30, especially at the read ends.
Per Sequence GC Content - GC content remained consistent and within the expected range, indicating no biases were introduced during trimming.
Adapter Content - Adapter sequences were successfully removed, confirming the effectiveness of the trimming process.
Sequence Duplication Levels - Sequence duplication levels decreased, indicating reduced PCR amplification biases and enhancing data reliability.
Kmer Content - Reduction in overrepresented sequences was observed, reflecting the effective removal of contaminants and adapter sequences.

Trimming significantly improved RNA-seq data quality by removing adapter sequences and reducing sequence duplication levels. These enhancements are crucial for accurate and reliable downstream analyses, such as transcriptome assembly and differential gene expression studies. Rigorous preprocessing ensures high-quality RNA-seq data, underscoring its importance in RNA-seq workflows.



MultiQC report of per sequence quality scores after trimming

Cuffdiff output

The Cuffdiff analysis produced a Gene_exp.diff file containing detailed information about genes aligned to the reference genome. By applying a significance filter ($p < 0.005$), 42 genes were identified as significant hits. These genes were then subjected to pathway enrichment analysis to explore their functional roles and associated biological pathways.

Gene hits from the cuffdiff output

gene_id	gene	p_value	q_value	significant
ENSG00000177082	WDR73	0.00005	0.00128541	yes
ENSG00000160818	GPATCH4	0.00005	0.00128541	yes
ENSG00000213676	ATF6B	0.00015	0.00330298	yes
ENSG00000100949	RABGGTA	0.00005	0.00128541	yes
ENSG00000185989	RASA3	0.00035	0.00670668	yes
ENSG00000137959	IFI44L	0.00005	0.00128541	yes
ENSG00000179051	RCC2	0.00005	0.00128541	yes
ENSG00000250673	REELD1	0.00165	0.0220883	yes
ENSG00000245468	LINC02447	0.00005	0.00128541	yes
ENSG00000079459	FDFT1	0.00005	0.00128541	yes
ENSG00000103415	HMOX2	0.0004	0.00746185	yes
ENSG00000130487	KLHDC7B	0.00005	0.00128541	yes
ENSG00000158352	SHROOM4	0.0003	0.00599304	yes
ENSG00000140740	UQCRC2	0.00005	0.00128541	yes
ENSG00000160932	LY6E	0.00005	0.00128541	yes
ENSG00000170161	FAM88B	0.0038	0.04057	yes
ENSG00000041982	TNC	0.001	0.0151239	yes
ENSG00000184226	PCDH9	0.00005	0.00128541	yes
ENSG00000120885	CLU	0.00005	0.00128541	yes
ENSG00000183160	TMEM119	0.0001	0.00231028	yes
ENSG00000152463	OLAH	0.00005	0.00128541	yes
ENSG00000109819	PPARGC1A	0.00005	0.00128541	yes
ENSG00000146592	CREB5	0.0001	0.00231028	yes
ENSG00000165409	TSHR	0.00285	0.0327993	yes

ENSG00000287839	-	0.00005	0.00128541	yes
ENSG00000166173	LARP6	0.00285	0.0327993	yes
ENSG00000184009	ACTG1	0.00005	0.00128541	yes
ENSG00000092094	OSGEP	0.00005	0.00128541	yes
ENSG00000118705	RPN2	0.00005	0.00128541	yes
ENSG00000111843	TMEM14C	0.00005	0.00128541	yes
ENSG00000103429	BFAR	0.00005	0.00128541	yes
ENSG00000197006	METTL9	0.00045	0.00813991	yes
ENSG00000218336	TENM3	0.00005	0.00128541	yes
ENSG00000075790	BCAP29	0.0009	0.0140219	yes
ENSG00000177706	FAM20C	0.00005	0.00128541	yes
ENSG00000171298	GAA	0.00005	0.00128541	yes
ENSG00000089737	DDX24	0.00005	0.00128541	yes
ENSG00000196497	IPO4	0.00005	0.00128541	yes
ENSG00000023191	RNH1	0.00005	0.00128541	yes
ENSG00000196943	NOP9	0.00155	0.0210958	yes
ENSG00000155974	GRIP1	0.0008	0.0128116	yes
ENSG00000166896	ATP23	0.00445	0.0456847	yes

Pathway enrichment

From David

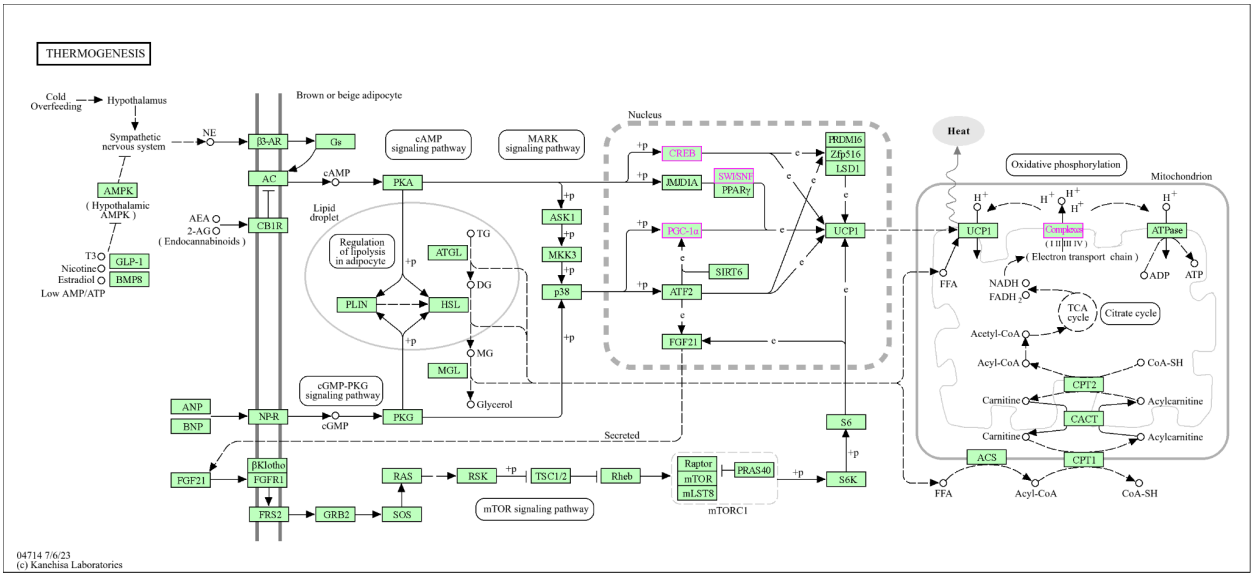
Category	Term
GOTERM_CC_DIRECT	GO:0005789~endoplasmic reticulum membrane
KEGG_PATHWAY	hsa04714:Thermogenesis
KEGG_PATHWAY	hsa04918:Thyroid hormone synthesis

From Shiny GO 0.8

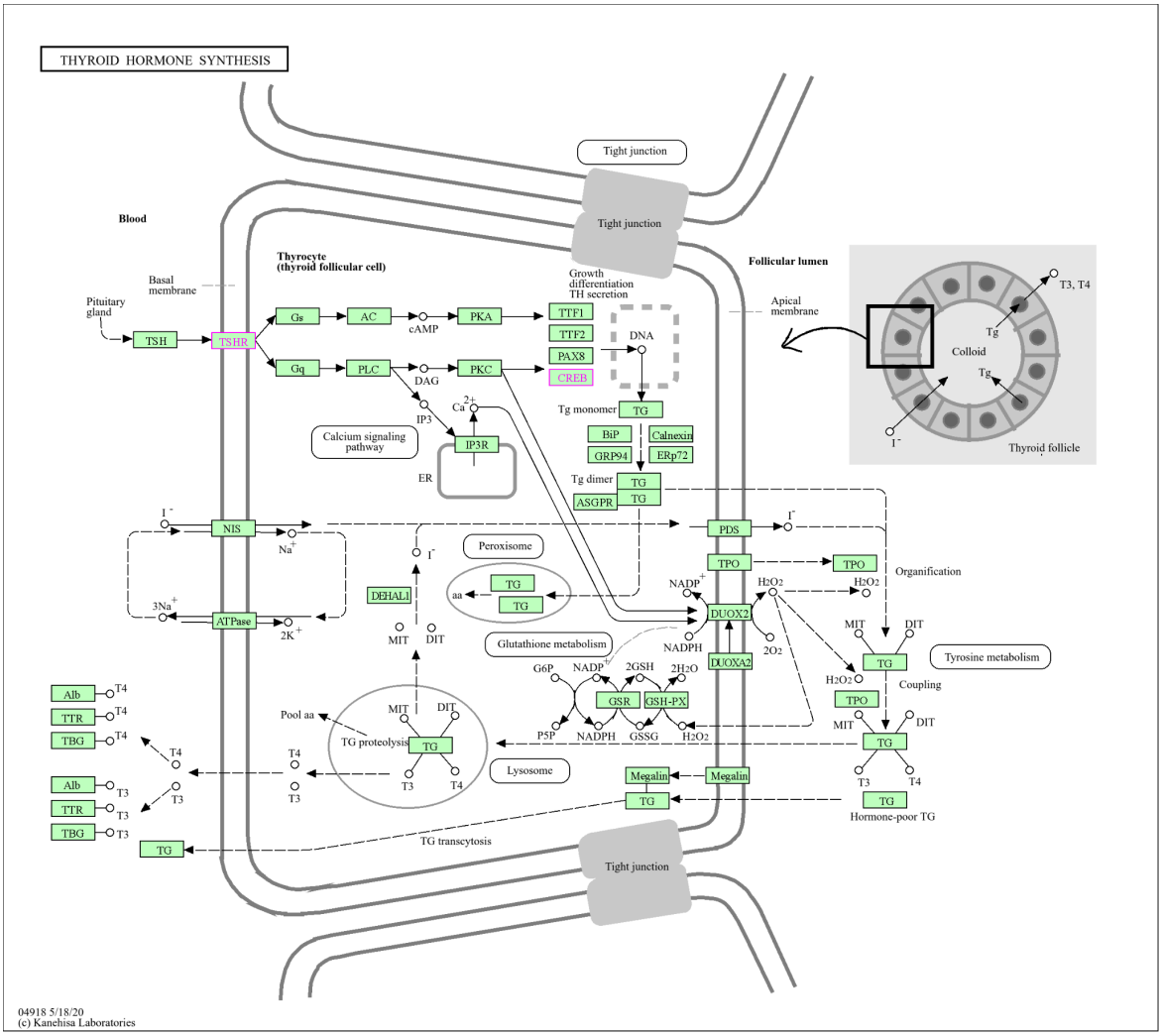
Pathway	Genes
Path:hsa04714 Thermogenesis	PPARGC1A, ACTG1, UQCRC2, CREB5
Path:hsa04211 Longevity regulating pathway	PPARGC1A, CREB5
Path:hsa04918 Thyroid hormone synthesis	TSHR, CREB5
Path:hsa05016 Huntington disease	PPARGC1A, UQCRC2, CREB5

Insights from NIH DAVID

Thermogenesis - This pathway involves processes of heat production essential for maintaining body temperature and energy balance. Among the genes in this pathway, three are particularly noteworthy. Additionally, two sub-pathways within thermogenesis are implicated in obesity, highlighting a potential link between the studied genes and metabolic disorders related to energy expenditure and fat accumulation.
Ref:- https://www.kegg.jp/network/nt06529+N01688+N01691

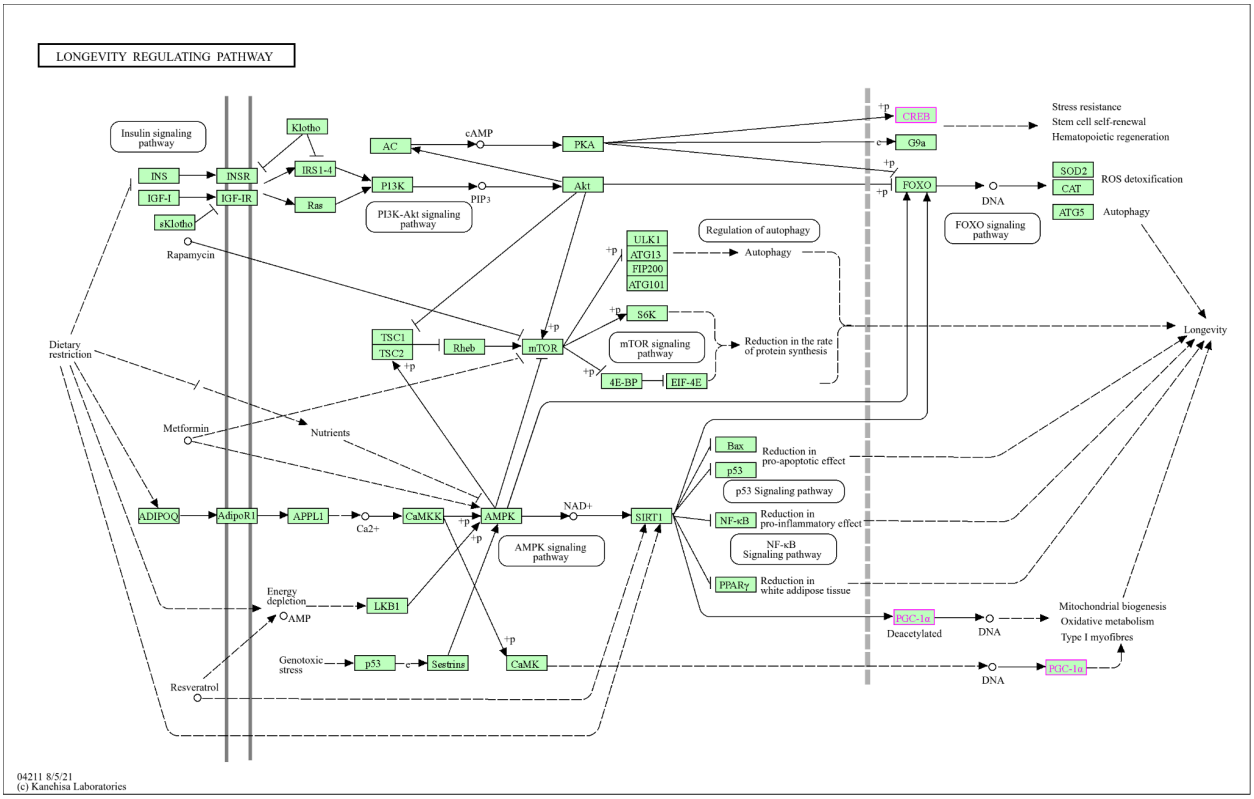


Thyroid hormone synthesis - Thyroid hormone synthesis encompasses several pathways, including the TRH-TSH-TH signaling pathway, where two significant genes are involved. Thyroid hormones are critical regulators of metabolism, growth, and development, and the enrichment in this pathway suggests that the studied genes may play a role in endocrine regulation and metabolic rate.
Ref:- https://www.kegg.jp/pathway/hsa04918



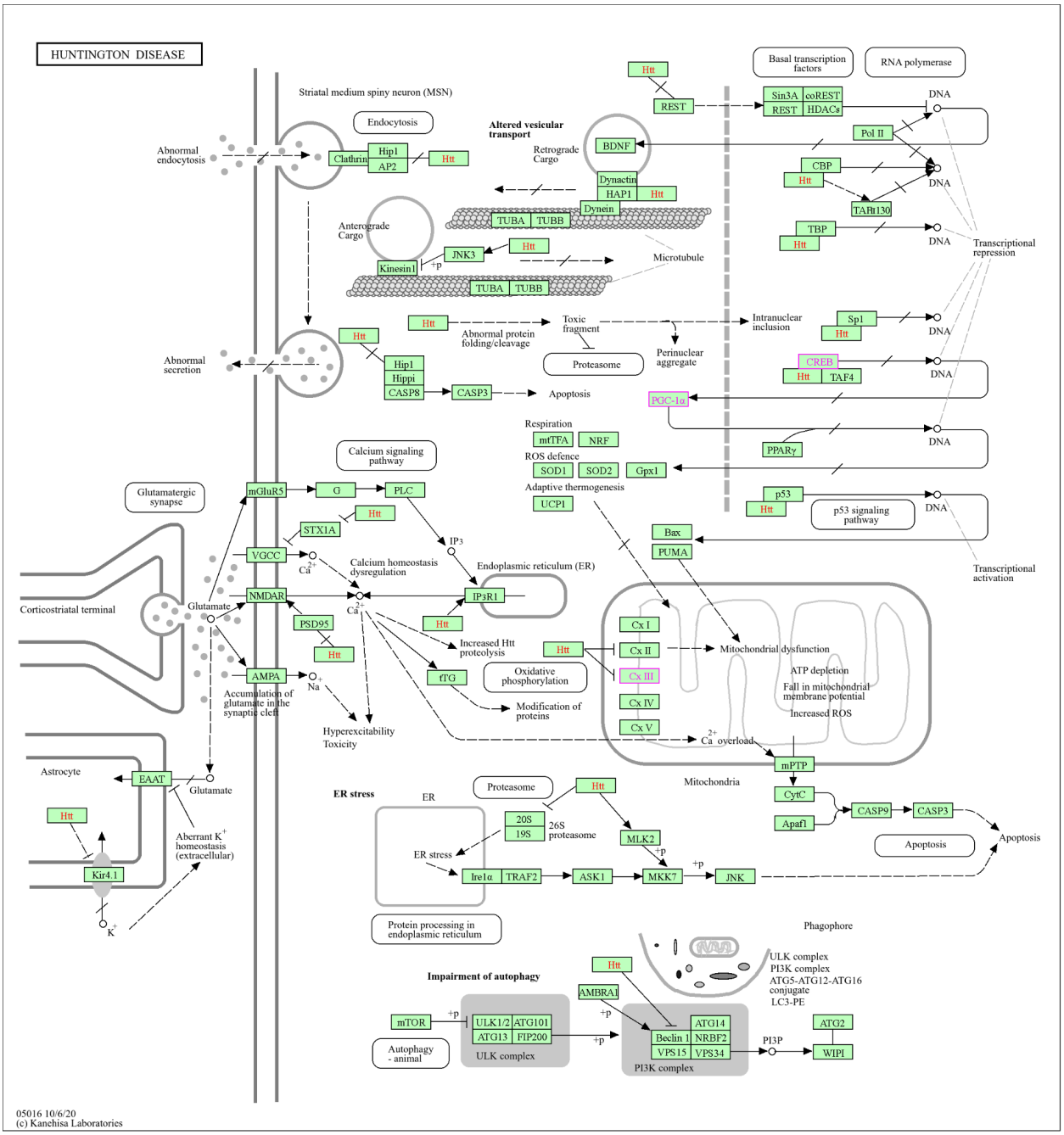
Insights from Shiny GO 0.8

Longevity regulating pathway - This pathway includes two significant genes and is crucial for understanding the genetic factors that influence lifespan and aging. The involvement of these genes in longevity regulation points to potential targets for interventions aimed at extending healthy life span and mitigating age-related diseases.
Ref:- https://www.kegg.jp/entry/hsa04211



Huntingtin Disease - Emerging evidence suggests that the mutant huntingtin protein, implicated in HD, may interfere with insulin signaling pathways, potentially predisposing HD patients to insulin resistance—a hallmark of T2D. Both T2D and HD share mitochondrial dysfunction as a common pathway. In T2D, mitochondrial dysfunction contributes to insulin resistance and β-cell failure, while in HD, it leads to neuronal cell death and neurodegeneration. This shared pathway underscores the critical role of mitochondrial health in both conditions, suggesting that improving mitochondrial function could be a therapeutic strategy for both T2D and HD.

Ref:- https://www.kegg.jp/entry/hsa05016



Conclusion

Using a transcriptomic approach, this project aimed to identify genes and metabolic pathways associated with Type 2 Diabetes (T2D) and Obesity in the context of betel nut chewing. Datasets from the GEO database, involving human monocyte cell lines exposed to Arecoline and MNPA (betel nut derivatives), were analyzed to uncover the genetic and metabolic disruptions caused by these compounds.

The pathway enrichment analysis revealed significant findings. The endoplasmic reticulum (ER) membrane pathway (GO:0005789) was enriched, highlighting the ER's crucial role in protein synthesis, folding, and lipid metabolism. This suggests that disruptions in ER function may contribute to metabolic dysregulation in T2D and Obesity.

The thermogenesis pathway (KEGG hsa04714) was also enriched, containing three notable genes. This pathway, essential for body temperature and energy balance, includes sub-pathways linked to obesity, indicating that betel nut constituents might influence metabolic rates and fat accumulation.

Additionally, the thyroid hormone synthesis pathway (KEGG hsa04918) was significant, particularly the TRH-TSH-TH signaling pathway involving two genes. This underscores the role of thyroid hormones in regulating metabolism and growth, reinforcing the hypothesis that betel nut derivatives may alter endocrine functions, contributing to metabolic diseases.

The longevity regulating pathway revealed two critical genes, suggesting potential impacts on aging and lifespan regulation. An intriguing connection between Huntington's Disease (HD) and T2D was also found. The mutant huntingtin protein, associated with HD, may interfere with insulin signaling, predisposing individuals to insulin resistance, a hallmark of T2D. Both conditions share mitochondrial dysfunction as a common pathogenic pathway, emphasizing the importance of mitochondrial health in both neuronal and metabolic integrity.

In conclusion, this project identified and validated several genes and their associated metabolic pathways potentially implicated in the development of T2D and Obesity due to betel nut chewing. These findings provide a deeper understanding of the molecular mechanisms at play and offer a valuable framework for future research to explore targeted therapeutic interventions, addressing the public health challenges posed by T2D and Obesity in populations with high rates of betel nut consumption.

Discussions

Exploring the influence of betel nut constituents on energy metabolism regulation, as indicated by the enrichment of the thermogenesis pathway, may offer insights into mechanisms underlying fat accumulation and metabolic rate modulation. This research could elucidate whether these compounds directly impact metabolic processes, potentially informing strategies to mitigate obesity-related health risks in betel nut-consuming populations.

Furthermore, the significant involvement of the thyroid hormone synthesis pathway suggests a broader impact on endocrine functions beyond ER and metabolic processes. Investigating the precise mechanisms through which betel nut derivatives modulate thyroid hormone signaling could provide insights into their role in metabolic dysregulation and offer avenues for therapeutic intervention targeting hormone imbalances associated with T2D and Obesity.

Additionally, the identification of genes associated with longevity pathways and their regulation hints at potential effects on aging processes in betel nut consumers. Understanding these mechanisms could lead to interventions that address accelerated aging or promote healthy aging in populations affected by betel nut use.

Finally, the intriguing connection between Huntington's Disease (HD) and T2D underscores the importance of mitochondrial health in both neuronal and metabolic integrity. Future studies could explore how mitochondrial dysfunction, exacerbated by betel nut constituents, contributes to insulin resistance and metabolic diseases. Investigating shared pathways between HD and T2D may reveal common therapeutic strategies targeting mitochondrial function to improve outcomes in both conditions.

In conclusion, the insights gained from this transcriptomic study provide a solid foundation for future research endeavors aimed at unraveling the complexities of T2D and Obesity exacerbated by betel nut chewing. By focusing on the identified pathways and genes, future studies can explore targeted therapeutic approaches to mitigate the health impacts observed in populations with high rates of betel nut consumption. Additionally, further investigation into the interplay between betel nut derivatives and mitochondrial function may offer promising avenues for developing interventions that address both metabolic and neurodegenerative disorders associated with these compounds.

References:

1) Shirleny R Cardosa, B. William Ogunkolade, Rob Lowe, Emanuel Savage, Charles A Mein, Barbara J Boucher, and Graham A Hitman. *Areca catechu*-(Betel-nut)-induced whole transcriptome changes in a human monocyte cell line that may have relevance to diabetes and obesity; a pilot study. BMC Endocr Disord. Lancet Diabetes Endocrinol 2021.
2) Prof Andrew P Hills, Prof Ross Arena, Prof Kamlesh Khunti, Prof Chittaranjan Sakerlal Yajnik, Ranil Jayawardena, Prof Christiani Jeyakumar Henry, Steven J Street, Mario J Soares, Prof Anoop Misra. Epidemiology and determinants of type 2 diabetes in south Asia. October 01, 2018.
3) Tung, TH., Chiu, YH., Chen, LS. et al. A population-based study of the association between areca nut chewing and Type 2 diabetes mellitus in men (Keelung Community-based Integrated Screening programme No. 2). Diabetologia 47, 1776–1781, 2004.
4) Inagaki T, Sakai J, Kajimura S. Transcriptional and epigenetic control of brown and beige adipose cell fate and function. Nat Rev Mol Cell Biol. 2016 Aug;17(8):480-95. doi: 10.1038/nrm.2016.62. Epub 2016 Jun 2. Erratum in: Nat Rev Mol Cell Biol. 2017 Aug;18(8):527.
5) Frühbeck G, Méndez-Giménez L, Fernández-Formoso JA, Fernández S, Rodríguez A. Regulation of adipocyte lipolysis. Nutr Res Rev. 2014 Jun;27(1):63-93.
6) Hondares E, Iglesias R, Giralt A, Gonzalez FJ, Giralt M, Mampel T, Villarroya F. Thermogenic activation induces FGF21 expression and release in brown adipose tissue. J Biol Chem. 2011 Apr 15;286(15):12983-90.
7) Yao L, Cui X, Chen Q, Yang X, Fang F, Zhang J, Liu G, Jin W, Chang Y. Cold-Inducible SIRT6 Regulates Thermogenesis of Brown and Beige Fat. Cell Rep. 2017 Jul 18;20(3):641-654.
8) Ahad F, Ganie SA. Iodine, Iodine metabolism and Iodine deficiency disorders revisited. Indian J Endocrinol Metab. 2010 Jan;14(1):13-7.
9) Marinò M, McCluskey RT. Role of thyroglobulin endocytic pathways in the control of thyroid hormone release. Am J Physiol Cell Physiol. 2000 Nov;279(5):C1295-306.
10) Barzilai N, Huffman DM, Muzumdar RH, Bartke A. The critical role of metabolic pathways in aging. Diabetes. 2012 Jun;61(6):1315-22.
11) Morris BJ, Willcox DC, Donlon TA, Willcox BJ. FOXO3: A Major Gene for Human Longevity--A Mini-Review. Gerontology. 2015;61(6):515-25.
12) Longo VD, Lieber MR, Vijg J. Turning anti-ageing genes against cancer. Nat Rev Mol Cell Biol. 2008 Nov;9(11):903-10.