Identification of Key Pathways and Genes in Obesity Using Bioinformatics Analysis and Molecular Docking Studies

Harish Joshi†, Basavaraj Vastrad‡, Nidhi Joshi§, Chanabasayya Vastrad¶, Anandkumar Tengli© and Iranna Kotturshetti†

† Department of Endocrinology, Endocrine and Diabetes Care Center, Hubballi, India, ‡ Department of Biochemistry, Basaveshwar College of Pharmacy, Gadag, India, § Department of Medicine, Dr. D. Y. Patil Medical College, Kolhapur, India, ¶ Bioinformatics, Chanabasava Nilaya, Bharthinagar, Dharwad, India, © Department of Pharmaceutical Chemistry, JSS College of Pharmacy, Mysuru and JSS Academy of Higher Education & Research, Mysuru, India, † Department of Ayurveda, Rajiv Gandhi Education Society’s Ayurvedic Medical College, Ron, India

Obesity is an excess accumulation of body fat. Its progression rate has remained high in recent years. Therefore, the aim of this study was to diagnose important differentially expressed genes (DEGs) associated in its development, which may be used as novel biomarkers or potential therapeutic targets for obesity. The gene expression profile of EMTAB-6728 was downloaded from the database. After screening DEGs in each ArrayExpress dataset, we further used the robust rank aggregation method to diagnose 876 significant DEGs including 438 up regulated and 438 down regulated genes. Functional enrichment analysis was performed. These DEGs were shown to be significantly enriched in different obesity related pathways and GO functions. Then protein–protein interaction network, target genes - miRNA regulatory network and target genes - TF regulatory network were constructed and analyzed. The module analysis was performed based on the whole PPI network. We finally filtered out STAT3, CORO1C, SERPINH1, MVP, ITGB5, PCM1, SIRT1, EEF1G, PTEN and RPS2 hub genes. Hub genes were validated by ICH analysis, receiver operating curve (ROC) analysis and RT-PCR. Finally a molecular docking study was performed to find small drug molecules. The robust DEGs linked with the development of obesity were screened through the expression profile, and integrated bioinformatics analysis was conducted. Our study provides reliable molecular biomarkers for screening and diagnosis, prognosis as well as novel therapeutic targets for obesity.

Keywords: adiposities, obesity, differentially expressed genes, modules, protein–protein interaction network

INTRODUCTION

Obesity has long been part of the larger metabolic disorder and affects a large proportion of the global population particularly in the Western World (1). Obesity is diagnosed on the basis of body mass index (1). Obesity occurs in children age between 5 to 19 years as well as more common in women than in men (2). Countless surveys have proved that obesity is an key risk factor for heart
disease (3), hyperlipidaemia (4), hyperinsulinaemia (5), hypertension (6), atherosclerosis (7), insulin resistance (8) and cancer (9). Important candidate genes and relevant signaling pathways linked with obesity remains largely unknown. As a result, seek of an earlier diagnosis and better prognosis, deeper understanding of genetic and molecular mechanisms about obesity is necessary.

Previous reports demonstrate that many genes and signaling pathways participate in obesity. Polymorphisms in UCP2 and UCP3 were responsible for development of obesity (10). TNFα and lipoprotein lipase were important for advancement of obesity (11). SLC6A14 (12) and JHDM2A (13) were lined with pathogenesis of obesity. Human salivary (AMY1) and pancreatic (AMY2) amylase genes were diagnosed with growth of obesity (14). Signaling pathways such as inflammatory signaling pathway (15), TLR4 signaling pathway (16), calcineurin-dependent signaling pathways (17), mTOR Complex1–S6K1 signaling pathway (18) and leptin-signaling pathway (19) were important for development of obesity. Therefore, it is meaningful to explore the precise molecular mechanisms involved in obesity and thus find a valid diagnostic way and generate an advance therapeutic strategy.

In present trends, the application of high-throughput analysis in gene expression profiling is becoming more valuable in clinical and medical research (20), molecular classification (21), prognosis prediction (22), diagnoses (23) and new targeted drug discovery (24). In this study, the original microarray data (E-MTAB-6728) was downloaded from ArrayExpress database (https://www.ebi.ac.uk/) and analyzed to get differently expressed genes (DEGs) between obesity persons and lean persons (normal controls). Subsequently, gene ontology (GO), pathway enrichment analysis, protein–protein interaction network construction and analysis, module analysis, target gene - miRNA interaction network construction and analysis, and target gene - TF interaction network construction and analysis to discover the key genes and pathways closely related to obesity. Finally, selected hub genes were validated by immunohistochemical (IHC) analysis, receiver operating characteristic curve (ROC) analysis and RT-PCR. This current investigation aimed at using bioinformatics tools to predict the key pathways and genes in obesity that can hold a value for target based therapeutic means.

**MATERIALS AND METHODS**

**Microarray Data**

The microarray expression profile of E-MTAB-6728 was downloaded from ArrayExpress (https://www.ebi.ac.uk/). E-MTAB-6728 was based on A-MEXP-1171 - Illumina HumanHT-12 v3.0 Expression BeadChip and was submitted by Bjune et al. (25). The E-MTAB-6728 dataset about expression of genes from obesity persons compared to lean persons (normal controls). There are twenty-four samples including twelve obesity persons and lean persons (normal controls). The overall design of the experiment was microarray analysis of adiposities from obese patients versus adipocytes from lean persons (controls).

**Identification of DEGs**

The raw data files were acquired for the analysis as IDAT files (illumina platform) forms and were converted into gene symbols and then processed to background correction and quantile data normalization using the effective multiarray average algorithm in the beadarray package (26). The analysis was carried out via R software (version 3.5.2). Hierarchical clustering analysis was applied to categorize the samples into two groups with similar expression patterns in obesity persons and lean persons (normal controls). The paired Student’s t-test based on the Limma package in R bioconductor was used to diagnose DEGs between two experimental conditions (27). Multiple testing corrections were performed by the Benjamini–Hochberg method (28). Then, the Log2 Fold change (log2FC) was determined. We selected up regulated DEGs with | log2 FC | > 0.524 and FDR < 0.05, and down regulated DEGs with | log2 FC | < -0.394 and FDR < 0.05 were considered as the cutoff values.

**Pathway and Gene Ontology (GO) Enrichment Analysis of DEGs**

The BIOCYC, Kyoto Encyclopedia of Genes and Genomes (KEGG), REACTOME, Pathway Interaction Database (PID), GenMAPP, MSigDB C2 BIOCARTA, PantherDB, Pathway Ontology and Small Molecule Pathway Database (SMDB) databases are a knowledge base for systematic analysis, annotation, and visualization of gene functions. The GO database can add functional classification for genomic data, including categories of biological processes (BP), cellular component (CC), and molecular function (MF). GO analysis is a prevalent genes and gene products annotating approach. ToppCluster (https://toppcluster.cchmc.org/) (29) is an online tool for gene functional classification, which is a key foundation for high-throughput gene analysis to understand the biological importance of genes. In the current investigation, in order to analyze the functions of DEGs, Pathway and GO enrichment analysis were conducted using the ToppCluster online tool; p<0.05 was set as the cutoff point.

**Integration of PPI Network and Module Analysis**

The mentha (https://mentha.uniroma2.it/) (30) is a biological database designed to predict protein-protein interaction (PPI) information. The DEGs were mapped to STRING to evaluate the interactive relationships, with a confidence score >0.9 defined as significant. Then, integration of protein–protein interaction (PPI) network was visualized using cytoscape software (version 3.8.2) (http://www.cytoscape.org/) (31). The plug-in Network Analyzer identified hub genes based on mathematical calculation methods such as node degree (32), betweenness (33), stress (34) and closeness (35) the number of genes within centrality mathematical calculation methods were represented the significance of the disorder. The PEWCC1 was applied to screen modules of PPI network with degree cutoff = 2, node score cutoff = 0.2, k-core = 2, and max. depth = 100 (36). The functional enrichment analysis in the module was performed by ToppCluster.
Construction of Target Genes - miRNA Regulatory Network

MiRNA of target genes were explored combined with the human miRNA information (miRNet database, (https://www.mirnet.ca/)) (37), recorded using TarBase, miRTarBase, miRecords, miR2Disease, HMDD, SM2miR, Phenomir, PharmacomiR, Epimir and starBase databases, and visualized using the Cytoscape software (31).

Construction of Target Genes - TF Regulatory Network

TFs of hub genes were explored combined with the human TF information (NetworkAnalyst database, http://www.networkanalyst.ca) (38), recorded using ENCODE database, and visualized using the Cytoscape software (31).

Validation of Hub Genes

Immunohistochemical (IHC) analysis of adipose tissues was performed utilizing human protein atlas (www.proteinatlas.org) (39). ROC analysis was performed using pROC package (40) in R. ROC analyses were estimated for diagnostic value of hub genes. When the AUC value was > 0.7, the hub genes were considered to be capable of distinguishing obesity persons from normal lean with excellent specificity and sensitivity.

Detection of the mRNA Expression of the Hub Genes by RT-PCR

D12 (ATCC CRL-3280) cell line for obesity and D16 (ATCC CRL-3281) cell line a normal control were purchased from the American Type Culture Collection (ATCC) (Maryland, USA). D12 cells were cultured in Dulbecco’s modified Eagle’s medium (DMEM) F12 medium, which contains 10% fetal bovine serum. D16 cells were cultured in Dulbecco’s modified Eagle’s medium (DMEM) F12 medium, which contains 10% fetal bovine serum. The culture temperature is 37°C and CO2 concentration is 5%. Total cellular RNA was extracted from cell culture with 1 ml TRI Reagent® (Sigma, USA). Reverse transcription cDNA kit (Thermo Fisher Scientific, Waltham, MA, USA) and random primers were used to synthesize cDNA. RT-PCR was performed using QuantStudio 7 Flex real-time PCR system (Thermo Fisher Scientific, Waltham, MA, USA). The conditions for RT-PCR amplification were as follows: 95°C for 120 seconds followed by 40 cycles of 95°C for 15 seconds, annealing temperature for 45 seconds. Each sample was run in triplicate. Relative expression level for each target gene was normalized by the Ct value of β-actin (internal control) using a 2^−ΔΔCT relative quantification method (41). The primer pairs used in the experiments are listed in Supplementary Table 1.

Molecular Docking Studies

The module SYBYL-X 2.0 perpetual software was used for Sulflex-Docking of the designed molecules. The molecules were sketched by using ChemDraw Software and imported and saved in sdf format using Openbabelfree software. The one co-crystallized protein from each of ERBB2, STAT3 and HSPAB8 were selected for docking studies. The protein structures of ERBB2, STAT3 and HSPAB8 of PDB code 1MFL, 5OOW and 3CWG was retrieved from Protein Data Bank (42–44). Together with the TRIPos force field, Gasteiger-Huckel (GH) charges were added to all designed molecules and the standard ant-obesity drug Orlistat, for the structure optimization process. In addition, energy minimization was carried out using MMFF94s and MMFF94 algorithm process. Protein processing was carried out after the incorporation of protein. The co-crystallized ligand and all water molecules were removed from the crystal structure; more hydrogen’s were added and the side chain was set. TRIPos force field was used for the minimization of structure. The designed molecules interaction efficiency with the receptor was represented by the Sulflex-Dock score in kcal/mol units. The interaction between the protein and the ligand, the best pose was incorporated into the molecular area. The visualization of ligand interaction with receptor is done by using discovery studio visualizer.

RESULTS

Data Normalization

Each array was normalized (centered) by quantile data normalization using the beadarray package in R bioconductor. As shown in Figures 1A, B, raw expression data were normalized after preprocessing; median-centered values demonstrated that the data were normalized and thus it was possible to cross-compare between obesity persons and lean persons (normal controls).

Identification of DEGs Between Obese Patients and Lean Persons

To preliminarily understand the mechanism contributing to the obesity, 24 patients [12 obesity persons and 12 lean persons (normal controls)] were selected for subsequent analysis. Based on the analysis, a total of 876 DEG compose of 438 genes had been expressed highly and about 438 genes had been shown to decrease expression in obesity and are listed in Supplementary Table 2. The FDR <0.05 was as a threshold value. Heat map is shown in Figure 2. Volcano plot for DEGs is shown in Figure 3.

Pathway and Gene Ontology (GO) Enrichment Analysis of DEGs

To further investigate the biologic functions and mechanisms of the DEGs, pathway and GO enrichment analyses were performed using ToppCluster tool. Pathway enrichment analysis revealed that the up regulated genes were mainly enriched in thyroid hormone metabolism II (via conjugation and/or degradation), ECM-receptor interaction, IL6-mediated signaling events, collagen formation, C21 steroid hormone metabolism, genes encoding collagen proteins, integrin signalling pathway, hypertension and suprofen pathway, and are listed in Supplementary Table 3. Similarly, down regulated genes were mainly enriched in superpathway of methionine degradation, ribosome, FoxO family signaling, eukaryotic translation elongation, propanoate metabolism, CDK regulation of DNA replication, p38 MAPK pathway, glycin, serine and threonine metabolic, and glycine, serine and...
FIGURE 1 | Box plots of the gene expression data before (A) and after normalization (B). Horizontal axis represents the sample symbol and the vertical axis represents the gene expression values. The black line in the box plot represents the median value of gene expression. (A1-A12 = adipocytes from lean persons; B1-B12 = adipocytes from obese patients).

FIGURE 2 | Heat map of differentially expressed genes.

FIGURE 3 | Volcano plot of differentially expressed genes. Genes with a significant change of more than two-fold were selected.
threonine metabolism, and are listed in Supplementary Table 4. GO analysis results showed that up regulated genes were significantly enriched in blood vessel morphogenesis, extracellular matrix and growth factor binding, and are listed in Supplementary Table 5. Similarly, down regulated genes were mainly enriched in organic acid biosynthetic process, cytosolic small ribosomal subunit and structural constituent of ribosome, and are listed in Supplementary Table 6.

Integration of PPI Network and Module Analysis

The PPI network of up regulated genes consisted of 7271 nodes and 16270 edges (Figure 4) and down regulated genes consisted of 7276 nodes and 19862 edges (Figure 5) constructed in the mentha database and visualized using Cytoscape software. Based on the mentha database, the DEGs with the highest PPI scores identified by the four centrality methods are shown in Supplementary Table 7. There are 5 up regulated genes selected as hub genes, such as HSPA8, HSPA5, YWHAH, STAT3 and ERBB2, and 5 down regulated genes selected as hub genes, such as ESR1, ARRB1, CSNK2A2, RBBP4 and NR3C1. A significant module was obtained from PPI network of DEGs using PEWCC1, including module 1 contains 49 nodes and 99 edges (Figure 6A) and module 2 contains 66 nodes and 754 edges (Figure 6B). Functional enrichment analysis revealed that genes in these modules were mainly involved in PI3K-Akt signaling pathway, regulation of nuclear SMAD2/3 signaling, ribosome, eukaryotic translation elongation, metabolism of amino acids and derivatives, disease,
cellular amide metabolic process, establishment of protein localization to endoplasmic reticulum, monocarboxylic acid biosynthetic process translation, translational initiation, macromolecule catabolic process and cytosolic small ribosomal subunit.

**Construction of Target Genes - miRNA Regulatory Network**

To further understand the regulatory network between miRNAs and target genes, through miRNet database were constructed by Cytoscape. As shown in Figure 7, the miRNA-regulated network with 2613 nodes (miRNA: 2261; target gene: 352) and 17260 edges was obtained for up regulated target genes and Figure 8, the miRNA-regulated network with 2685 nodes (miRNA: 2327; target gene: 358) and 19827 edges was obtained for down regulated target genes. Different target genes regulated by miRNAs are shown in Supplementary Table 8. SOD2 had been predicted to regulate 257 miRNAs (ex; hsa-mir-3144-3p), CCND1 had been predicted to regulate 251 miRNAs (ex; hsa-mir-7706), TUBB2A had been predicted to regulate 193 miRNAs (ex; hsa-mir-5692c), CCND2 had been predicted to regulate 179 miRNAs (ex; hsa-mir-7162-3p), TMEM189 had been predicted to regulate 146 miRNAs (ex; hsa-mir-548z), BTG2 had been predicted to regulate 247 miRNAs (ex; hsa-mir-6075), TXNIP had been predicted to regulate 228 miRNAs (ex; hsa-mir-3194-3p), MED28 had been predicted to regulate 203 miRNAs (ex; hsa-mir-6861-5p), CNBP had been predicted to regulate 197 miRNAs (ex; hsa-mir-4651) and MKNK2 had been predicted to regulate 195 miRNAs (ex; hsa-mir-3650).

**Construction of Target Genes - TF Regulatory Network**

To further understand the regulatory network between TFs and target genes, through NetworkAnalyst database were constructed by Cytoscape. As shown in Figure 9, the TF-regulated network with 629 nodes (TF: 336; Gene: 293) and 6293 edges was obtained for up regulated target genes and Figure 10, the TF-regulated network with 2685 nodes (TF: 342; Gene: 299) and 8597 edges was obtained for down regulated target genes. Different target genes regulated by TFs are shown in Supplementary Table 9. YWHAH had been predicted to regulate 70 TFs (ex; MAZ), LYZ had been predicted to regulate 62 TFs (ex; TFDP1), HP had been predicted to regulate 60 TFs (ex; KLF9), TRAM2 had been predicted to regulate 54 TFs (ex; KLF16), CCND1 had been predicted to regulate 51 TFs...
(ex; EZH2), EFNA1 had been predicted to regulate 91 TFs (ex; TFDP1), MEDI6 had been predicted to regulate 85 TFs (ex; MAZ), RWDD2A had been predicted to regulate 82 TFs (ex; KDM5B), ADD3 had been predicted to regulate 82 TFs (ex; SAP30) and AIP had been predicted to regulate 82 TFs (ex; PHF8).

Validation of Hub Genes

Immunohistochemical analysis demonstrated that the expression of STAT3, CORO1C, SERPINH1, MVP and ITGB5 were highly expressed in adipose tissues, whereas PCM1, SIRT1, EEF1G, PTEN and RPS2 were low expressed in adipose tissue (Figure 11I) and Box plots is showed in Figure 11II. Validated by ROC curves, we found that 10 hub genes had high sensitivity and specificity, including STAT3 (0.951), CORO1C (0.799), SERPINH1 (0.924), MVP (0.938), ITGB5 (0.938), PCM1 (0.826), SIRT1 (0.799), EEF1G (0.913), PTEN (0.833) and RPS2 (0.840) (Figure 12). The 10 hub genes might be biomarkers of obesity and have positive implications for early medical intervention of the disease.

Detection of the mRNA Expression of the Hub Genes by RT-PCR

The adipocytes were removed to detect the mRNA expression levels of hub genes in the PPI network, including STAT3, CORO1C, SERPINH1, MVP, ITGB5, PCM1, SIRT1, EEF1G, PTEN and RPS2. It was found that the mRNA expression levels of STAT3, CORO1C, SERPINH1, MVP and ITGB5 were significantly increased in the obesity compared with the control group. Furthermore, the results illustrate that the mRNA expression levels of PCM1, SIRT1, EEF1G, PTEN and
RPS2 were significantly decreased in the obesity compared with the control group (Figure 13). Therefore, the RT-PCR results of the hub genes were consistent with the bioinformatics analysis.

**Molecular Docking Studies**

In the present research, the docking simulations are performed to identify the active site conformation and major interactions responsible for complex stability with the ligand receptor. Designed novel molecules containing four membered more sensitive β-lactam ring, the four membered and performed docking studies using Sybyl X 2.1 drug design software. Molecules containing β-lactam ring is designed which is easily reacting group Figure 14A, based on the structure of anti-obesity drug orlistat four membered ring Figure 14B, has potent pancreatic lipase inhibitory activity. The molecules were designed based on the structure of the standard anti-obesity drug orlistat. The one protein in each of three over expressed genes of ERBB2, its co-crystallized protein of PDB code 1MFL, HSPAB 8 its co-crystallized protein of PDB code 5OOW and STAT 3 its co-crystallized protein of PDB code of 3CWG respectively selected for docking studies. The investigation of designed molecules was performed to identify the potential molecule. The most of the designed molecules with respect to the standard anti-obesity drug orlistat, obtained C-score greater than 5. The C-score greater than 5 are said to be an active, among total of 32 designed molecules few molecules have excellent good binding energy (C-score) greater than 7 respectively.
FIGURE 12 | ROC curve validated the sensitivity, specificity of hub genes as a predictive biomarker for obesity prognosis. (A) STAT3 (B) CORO1C (C) SERPINH1 (D) MVP (E) ITGB5 (F) PCM1 (G) SIRT1 (H) EEF1G (I) PTEN (J) RPS2.
FIGURE 13 | Validation of hub genes by RT-PCR. (A) STAT3 (B) CORO1C (C) SERPINH1 (D) MVP (E) ITGB5 (F) PCM1 (G) SIRT1 (H) EEF1G (I) PTEN (J) RPS2.
The molecule ND4, FU5 and PF5 obtained score of 7.242, 7.659 and 7.842 with 1MFL and the molecules PM6, ND1, ND3, ND5, ND6, PF5 and PF6 obtained score of 7.5269, 7.6271, 8.0824, 7.6595, 7.0792 and 7.2659 with 3 CWG and the molecules PM4, PM6, ND1, ND5, ND6, PF4, and PF obtained good binding score of 7.1631, 8.8312, 7.3781, 7.9872, 7.9567, 7.0213 and 7.0386 with 5OOW respectively. The molecules found binding score 5-6 is PM1, PM2, PM3, PM4, PM5, PM6, PM7, PM8, ND1, ND2, ND3, ND5, ND6, ND7, ND8, FU1, FU2, FU3, FU4, FU7, FU8, PF1, PF2, PF3, PF4, PF6, PF7 and standard olistat (STD) with 1MFL and PM2, PM6, FU17, FU18, FU19, FU20, FU23, PF26, PF27, PF28 and PF32 with 3CWG, and PM1, PM2, PM3, PM5, PM7, PM8, ND2, ND3, ND4, ND7, ND8, FU1, FU2, FU3, FU4, FU5, FU6, FU7, FU8, PF1, PF2, PF3, PF6, PF7 and PF8 5OOW respectively. No molecules obtained binding score with less than 5 respectively; the values are depicted in Supplementary Table 10. The molecule PF5 has good binding score with all three proteins and ND1, ND3, ND5 and ND6 obtained good binding score with 3CWG and 5OOW. The molecule ND5 has highest binding score and is very close with standard olistat, the interaction with protein 5OOW and hydrogen bonding and other bonding interactions with amino acids are depicted by 3D (Figure 15) and 2D (Figure 16) images.

**DISCUSSION**

Due to the heterogeneity of obesity, obesity was still a disease with high rates of prevalence. This might be due to the scarcity of
valid biomarkers for detection of obesity and of valid treatment for obesity. Therefore, molecular mechanisms of obesity are necessary for scientists to find the treat and diagnosis method of obesity. Because of the fast advancement of bioinformatics analysis, it is more convenient to find out the genetic modification in obesity. Bioinformatics analysis enables us to explore the gene, the genetic change in obesity, which had been proved to be a better approach to identify novel biomarkers.

In our study, a total of 876 DEGs were diagnosed from gene expression dataset, consisting of 438 up regulated genes and 438 down regulated genes in obese persons compared to lean persons. Study showed that PTGDS (prostaglandin D2 synthase) (45), LBP (lipopolysaccharide binding protein) (46), EGFL6 (47), STAT3 (48) and HDAC9 (49) were closely associated with obesity. The expression level of CYP11A1 (50) and WNT11 (51) were linked to cancer progression, but these genes might be novel target for obesity. A previous study showed that expression of GPR146 played an important role in insulin resistance (52), but these genes might be novel target for obesity. Aberrant expression of RFX1 (53) and (54) are noticeable factors in the heart disease, but these genes might be novel target for obesity. CLDN1 expression predicted poor therapeutic outcomes of hypertension patients (55).

Functional enrichment analysis of DEGs was implemented. SULT1A1 (56), SULT1A2 (56), COL6A1 (57), COL6A2 (58), SOS1 (59), STAT1 (60), COL5A2 (61), RND3 (62), COL15A1 (63), CBS (cystathionine-beta-synthase) (64), MCM6 (65), TNFRSF12A (66), FMOD (fibromodulin) (67), TYMP (thymidine phosphorylase) (68), ALPL (alkaline phosphatase, biomineralization associated) (69), EFEMP1 (70), MFAP4 (71), SULT1A1 (72), SULT1A2 (72), COL11A1 (73), COL12A1 (74), SERPINH1 (75), RHOC (76), RPL14 (90), RPL29 (91), RPS12 (92), RPS15A (93), RPS2 (94), RPS27 (95), RPS3 (96), RBL2 (97), EEF2F (98), ACACA (acyl-CoA carboxylase alpha) (99), ORC2 (100), PHGDH (phosphoglycerate dehydrogenase) (101), SHMT1 (102), NRCAM (neuronal cell adhesion molecule) (103), NRP2 (104), RP30 (105), SRPX2 (106), THY1 (107), CD248 (108), CLEC3B (109), CST3 (110), CTBRC1 (111), GPC1 (112), ACS5 (113) and HSD17B12 (114) have been reported to be biomarkers of heart disease or play a vital role in its pathogenesis, but these genes might be novel target for obesity. Reports illustrate that SULT1A1 (115), THBS1 (116), TIMP1 (117), LOX (lysyl oxidase) (118), MMP9 (119), HSD11B1 (120), ITGEB2 (121), HMOX1 (122), SOD2 (123), AKR1C3 (124), MAT2B (125), FOXO1 (126), FOXO3 (127), SIRT1 (128), ACACB (acyl-CoA carboxylase beta) (129), ELK1 (130), MAP3K5 (131), CTH (cystathionine gamma-lyase) (132), AMOT (angiomotin) (133), CDDC80 (134), CXCL10 (135), ERBB2 (136), KLF4 (137), LEP (leptin) (138), MFGE8 (139), SLT2 (140), TNMD (tenomodulin) (141), ADAMTS5 (142), ELN (elastin) (143), HTRA1 (144), LUM (lumican) (145), MFA5 (146), IL1RN (147), ACADL (acyl-CoA dehydrogenase long chain) (148), AGT (angiotensinogen) (149), FADS1 (150), PDK4 (151), PER2 (152) and SLC27A2 (153) in obesity. CDKN1B was shown to be a potential predictor of advanced hyperinsulinemia (154), but this gene might be novel target for obesity. Reports illustrate that CXCL12 (155) and IGFBP6 (156) and ELOVL6 (157) were expressed in patients with insulin resistance, but these genes might be novel target for obesity.

Furthermore, by constructing PPI networks and modules, we identified some key genes that provide new insights for obesity diagnosis, prognosis, and drug target identification. Expression of the HSPA8 (158) and CKB (159) were correlated with disease grades of hypertension, but these genes might be novel target for obesity. Recent studies have proposed that TPTA (160), YWHAH (161), ESR1 (162), PTEN (163), IRAK1 (164), CYR61 (165) and ZBTB16 (166) are involved in obesity. Previous reports demonstrate that SPTAN1 (167), STEAP2 (168), NEK6 (169), ARRB1 (170), FBXO11 (171), UBR2 (172), JN56 (173), CDK4 (174), LMO2 (175), MZN (176), TAGLN2 (177), SRFSE (178), SAFB (179), SINA (180), TRIM24 (181) and AUTS2 (182) appears to be constitutively activated in cancer, but these genes might be novel target for obesity. CSN2KA2 expression might be regarded as an indicator of susceptibility to heart disease (183), but this gene might be novel target for obesity. COPG2, FBL, CSNK2B, PCMP, ZNF581, KHDRBS1, RBMX, RBBP4 and DCAF7 are novel biomarkers for obesity.

Target genes - miRNA regulatory network and target genes - TF regulatory network were constructed and analyzed. A previous study reported that CCND1 (184) and HP (185) were expressed in obesity. CCND2 (186) and TXNIP (187) are a potential marker for the detection and prognosis of insulin resistance, but these genes might be novel target for obesity. Recent research has revealed that BTG2 was expressed in obesity (188). Expression of MED8 (189) and EFN1 (190) might participate in cancer progression, but these genes might be novel target for obesity. TUBB2A,
TMEM189, CNBP, LYZ, TRAM2, MED16, RWDD2A, ADD3 and AIP are novel biomarkers for obesity.

However, this investigation had some limitations. Primarily, the mechanisms of several hub genes in the pathological process of obesity remain unclear, warranting needs further investigation. Moreover, the success of our small molecule drug compound screening in reducing obesity remains to be assessed.

In conclusion, in this study, we determined that STAT3, CORO1C, SERPINH1, MVP, ITGB5, PCMI, SIRT1, EEF1G, PTEN and RPS2 might be critical genes in the development and prognosis of obesity through bioinformatics analysis combined with validations. However, it is essential that further experiments are carried out and clinical data made available to confirm the results of our investigation and guide the discovery of future gene therapies against obesity.

**DATA AVAILABILITY STATEMENT**

The datasets supporting the conclusions of this article are available in the ArrayExpress (https://www.ebi.ac.uk/arrayexpress/) repository. [E-MTAB-6728](https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-6728/).

**REFERENCES**

1. Cole TJ, Bellizzi MC, Flegal KM, Dietz WH. Establishing a Standard Definition for Child Overweight and Obesity Worldwide: International Survey. BMJ (2000) 320(7244):1240–3. doi: 10.1136/bmj.320.7244.1240
2. Krotkiewski M, Björntorp P, Sjöström L, Smith U. Impact of Obesity on Metabolism in Men and Women. Importance of Regional Adipose Tissue Distribution. J Clin Invest (1983) 72(3):1150–62. doi: 10.1172/JCI11040
3. Freedman DS, Khan LK, Dietz WH, Srinivasan SR, Berenson GS. Relationship of Childhood Obesity to Coronary Heart Disease Risk Factors in Adulthood: The Bogalusa Heart Study. Pediatrics (2001) 108(3):712–8. doi: 10.1542/peds.108.3.712
4. Locsey L, Asztalos L, Kincses Z, Berczi C, Paragh G. The Importance of Obesity and Hyperlipidaemia in Patients With Renal Transplants. Int Urol Nephrol (1998) 30(6):767–75. doi: 10.1007/BF02564866
5. Becker S, Dowssus L, Kaaks R. Obesity Related Hyperinsulinaemia and Hyperglycaemia and Cancer Development. Arch Physiol Biochem (2009) 115(2):86–96. doi: 10.1080/13813450902878054
6. Rahmouni K, Correia ML, Haynes WG, Mark AL. Obesity-Associated Hypertension: New Insights Into Mechanisms. Hypertension (2005) 45(1):1–14. doi: 10.1161/01.HYP.0000151325.83008.b4
7. Lovren F, Teoh H, Verma S. Obesity and Atherosclerosis: Mechanistic Insights. Can J Cardiol (2015) 31(2):77–83. doi: 10.1016/j.cjca.2014.11.031
8. Steppan CM, Bailey ST, Bhat S, Brown EJ, Banerjee RR, Wright CM, et al. The Hormone Resistin Links Obesity to Diabetes. Nature (2001) 409(6818):307–12. doi: 10.1038/35053000
9. Calle EE, Kaaks R. Overweight, Obesity and Cancer: Epidemiological Evidence and Proposed Mechanisms. Nut Rev Cancer (2004) 4(8):579–91. doi: 10.1038/sj.nrc1408
10. Jia JJ, Zhang X, Ge CR, Jois M. The Polymorphisms of UCP2 and UCP3 Genes Associated With Fat Metabolism, Obesity and Diabetes. Obes Rev (2009) 10(5):519–26. doi: 10.1111/j.1467-789X.2009.00569.x
11. Kern PA. Potential Role of TNFalpha and Lipoprotein Lipase as Candidate Genes for Obesity. J Nutr (1997) 127(9):1917S–22S. doi: 10.1093/jn/127.9.1917S
12. Suviohtel E, Oksanen LJ, Ohman M, Cantor RM, Ridderstrale M, Tuomi T, et al. The SLC6A14 Gene Shows Evidence of Association With Obesity. J Clin Invest. (2003) 112(11):1762–72. doi: 10.1172/JCI17491

**AUTHOR CONTRIBUTIONS**

HJ: Methodology and validation. BV: Writing original draft, and review and editing. NJ: Software and resources. CV: Investigation and resources. AT: Formal analysis and validation. IK: Supervision and resources. All authors contributed to the article and approved the submitted version.

**ACKNOWLEDGEMENTS**

We thank Jan-Inge Bjune, Gunnar Mellgren and Simon N Dankel UNIVERSITY IN BERGEN, The Hormone Laboratory, 8.etg Lab Building, Jonas Lies road 91B, 5021 Bergen very much, the authors who deposited their microarray dataset, E-MTAB-6728, into the public Array Express database.

**SUPPLEMENTARY MATERIAL**

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fendo.2021.628907/full#supplementary-material
Joshi et al. Identification of Key Genes in Obesity

24. Jia L, Jia R, Li Y, Li X, Jia Q, Zhang H. LCK as a Potential Therapeutic Target for Acute Rejection After Kidney Transplantation: A Bioinformatics Clue. J Immunol Res (2018) 2018:6451298. doi: 10.1155/2018/6451298

25. Bijnne J, Haugen G, Guldbrandsen O, Nordbe OP, Nielsen H, Våge V, et al. IRX5 Regulates Adipocyte Amlyoid Precursor Protein and Mitochondrial Respiration in Obesity. Int J Obes (Lond) (2018) 43(11):2151–62. doi: 10.1038/s41366-018-0275-y

26. Dunning MJ, Smith ML, Ritchie ME, Tavaré S. Beadarray: R Classes and Methods for Illumina Bead-Based Data. Bioinformatics (2007) 23(16):2183–4. doi: 10.1093/bioinformatics/btm311

27. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. Limma Powers Differential Expression Analyses for RNA-Sequencing and Microarray Studies. Nucleic Acids Res (2015) 43(7):e47. doi: 10.1093/nar/gko037

28. Abbas A, Kong XB, Liu Z, Jing BY, Gao X. Automatic Peak Selection by a Benjamini-Hochberg-based Algorithm. PloS One (2013) 8(1):e53112. doi: 10.1371/journal.pone.0053112

29. Kaimal V, Bardes EE, Tabar SC, Jegga AG, Aronow BJ. ToppCluster: A Multiple Gene List Feature Analyzer for Comprehensive Enrichment Clustering and Network-Based Dissection of Biological Systems. Nucleic Acids Res (2010) 38(Web Server issue):w96–W102. doi: 10.1093/nar/gkq418

30. Calderone A, Castagnoli L, Cesareni G. Mentha: A Resource for Browsing Integrated Protein-Interaction Networks. BMC Research Notes (2010) 3(1):248–504. doi: 10.1101/gr.123930

31. Przulj N, Wigle DA, Jurisica I. Functional Topology in a Network of Protein Interactions. Bioinformatics (2004) 20(3):340–8. doi: 10.1093/bioinformatics/btg415

32. Nguyen TP, Liu WC, Jorda C, Almeida A, Zhao Q, Siegel RJ, et al. Increasingly Expressed in Human Obesity and Promotes Proliferation of Adipose Tissue-Derived Stromal Vascular Cells. Mol Cell Biochem (2010) 343(1–2):257–69. doi: 10.1007/s11010-010-0521-7

33. Veeramachaneni GK, Raj KK, Chalasani LM, Annamraju SK, Is B, Talluri VR. Shape Based Virtual Screening and Molecular Docking Towards Designing Novel Pancreatic Lipase Inhibitors. Bioinformation (2015) 11(12):335–42. doi: 10.6026/97320630011535

34. Urbanet R, Nguyen Dinh Cat A, Feraco A, Venteclf N, El Mogrhabi S, Siegerra-Ramos C, et al. Adipocyte Mineralcorticoid Receptor Activation Leads to Metabolic Syndrome and Induction of Prostaglandin D2 Synthase. Hypertension (2015) 66(1):149–57. doi: 10.1161/HYPERTENSIONAHA

35. Fadhal E, Gamieldien J, Mwambene EC. Protein Interaction Networks as a Tool for Understanding Human Infection. Frontiers in Endocrinology (2015) 6(52):1–17. doi: 10.3389/fendo.2015.00027

36. Zhou G, Soufan O, Ewald J, Hancock REW, Basu N, Xia J. NetworkAnalyst: A Web-Based Platform for Complex Multi-Omics Data Analysis. Nucleic Acids Res (2018) 46(W1):W102. doi: 10.1093/nar/gkq418

37. El-Saadi MW, Williams-Hart T, Salvatore BA, Mahdavian E. Use of in-Silico Assays to Characterize the ADMET Properties of Novel Peptide Inhibitors of IRF8 in Vascular Inflammation and Cardiovascular Disease: Diagnostic and Therapeutic Potential. Int Rev Immunol (2016) 35(5):434–54. doi: 10.3109/08830185.2015.1078519
Joshi et al. Identification of Key Genes in Obesity

61. Azuaje F, Zhang L, Jeanty C, Puhl SL, Rodius S, Wagner DR. Analysis of a Gene Co-Expression Network Establishes Robust Association Between Col5a2 and Ischemic Heart Disease. BMC Med Genomics (2013) 6:13. doi: 10.1186/1755-8794-6-13

62. Yue X, Yang X, Lin X, Yang T, Yi X, Dai Y, et al. Rnd3 Halopsin-Sufficient Mice are Predisposed to Hemodynamic Stress and Develop Apoptotic Cardiomyopathy With Heart Failure. Cell Death Dis (2014) 5:e1284. doi: 10.1038/cddis.2014.235

63. Connelly JJ, Cherepanova OA, Doss JF, Karaoli T, Lillard TS, Markunas CA, et al. Epigenetic Regulation of COL1A1 in Smooth Muscle Cell Replicative Aging and Atherosclerosis. Hum Mol Genet (2013) 22(25):5107–20. doi: 10.1093/hmg/ddt365

64. Karamat FA, Oudman I, Ris-Stalpers C, Anezonkú M, Pinto-Sietsma S, et al. Tnfrsf12a-Mediated High Glucose-Induced Cardiac Fibroblast Activation. J Mol Endocrinol (2013) 8(12):e82243. doi: 10.1530/JME-12-0194

65. Smith CE, Coltell O, Sorlí JV, Estruch R, Martínez-González MA, Salas-Salvadó J, et al. Associations of the MCM6-rs3754686 Proxy for Milk Intake in Mediterranean and American Populations With Cardiovascular Biomarkers, Disease and Mortality: Mendelian Randomization. Sci Rep (2016) 6:33188. doi: 10.1038/srep33188

66. Lyu M, Cui Y, Zhao T, Ning Z, Ren J, Jin X, et al. Tnfrsf12a-1Mediated Atherosclerosis Signaling and Inflammatory Response as a Common Protection Mechanism of Shuangmei Injection Against Both Myocardial and Cerebral Ischemia-Reperfusion Injuries. Front Pharmacol (2018) 9:312. doi: 10.3389/fphar.2018.00312

67. Shami A, Tengryd C, Asciutto G, Bengtsson E, Nilsson J, Hultgårdh-Nilsson L, et al. Association Of The MCM6-rs3754686 Proxy For Milk Intake In Mediterranean and American Populations With Cardiovascular Biomarkers, Disease and Mortality: Mendelian Randomization. Sci Rep (2016) 6:33188. doi: 10.1038/srep33188

68. Haarhaus M, Brandenburg V, Kalantar-Zadeh K, Stenvinkel P, Magnusson OW, et al. Thymidine Phosphorylase: A Potential New Target for Hypertension. Hypertension (2014) 63(1):68–73. doi: 10.1161/HYPERTENSIONAHA.113.01352

69. Smith CE, Coltell O, Sorlí JV, Estruch R, Martinez-Gonzalez MA, Salas-Salvado J, et al. Associations of the MCM6-rs3754686 Proxy for Milk Intake in Mediterranean and American Populations With Cardiovascular Biomarkers, Disease and Mortality: Mendelian Randomization. Sci Rep (2016) 6:33188. doi: 10.1038/srep33188

70. He H, Wang J, Yannie PJ, Kakiyama G, Korzun WJ, Ghosh S. Sterol Carrier Protein Stm2a-Rho GTPase Activity Regulates the Development of Hypertension Through Oxidative Stress-Dependent Vascular Inflammation. Am J Physiol Heart Circ Physiol (2013) 305(5):H756–68. doi: 10.1152/ajpheart.00373.2013

71. Andraweeha PH, Dekker GA, Thompson SD, North RA, McCowan LM, Roberts CT. A Functional Variant in ANGPT1 and the Risk of Pregnancies With Hypertensive Disorders and Small-for-Gestational-Age Infants. Mol Hum Reprod (2012) 18(6):325–32. doi: 10.1093/molehr/gar081

72. Li X, Zhang X, Leathers R, Makino A, Huang C, Parsa P, et al. Notch3 Signaling Promotes the Development of Pulmonary Arterial Hypertension. Nat Med (2009) 15(11):1289–97. doi: 10.1038/nm.2021

73. Usui T, Okada M, Hara Y, Yamawaki H. Eukaryotic Elongation Factor 2 Kinase Regulates the Development of Hypertension Through Oxidative Stress-Dependent Vascular Inflammation. Am J Physiol Heart Circ Physiol (2013) 305(5):H756–68. doi: 10.1152/ajpheart.00373.2013

74. Chen J, Zhao X, Wang H, Chen Y, Wang W, Zhou L, et al. Common Variants in TGFRB2 and miR-518 Genes Are Associated With Hypertension in the Chinese Population. Am J Hypertens (2014) 27(10):1268–76. doi: 10.1038/ajh/rapid047

75. Chamelevieva P, Besse A, Boucard N, Simon A, Leroux D, Pintel N, et al. Dag-1 Carcinoma Cell in Studying the Mechanisms of Progression and Therapeutic Resistance in Bladder Cancer. Eur Urol (2001) 39(3):343–8. doi: 10.1159/000052465

76. Flum M, Kleemann M, Schneider H, Wei B, Fischer S, Handrick R, et al. miR-217:5p Induces Apoptosis by Directly Targeting PRKCI, Bag3, ITGAV and MAP1K in Colorectal Cancer Cells. J Cell Commun Signal (2018) 12 (2):451–66. doi: 10.1007/s12079-017-0410-x

77. Kamikawaji K, Seki N, Watanabe M, Matakai T, Takagi K, et al. Regulation of LOXL2 and SERPINH1 by Antitumor microRNA-29a in Lung Cancer With Idiopathic Pulmonary Fibrosis. J Cell Sci (2016) 6(3):68. doi: 10.1093/hmg/ddt365

78. Konishi H, Okuda A, Ohno Y, Kihara A. Characterization of HACD1 K64Q Mutant Found in Arrhythmogenic Right Ventricular Dysplasia Patients. J Biochem (2010) 148(5):617–23. doi: 10.1093/jb/mvq092

79. Liu BL, Sun KK, Zong ZH, Chen S, Zhao Y. MicroRNA-372 Inhibits Endometrial Carcinoma Development by Targeting the Expression of the Ras Homolog Gene Family Member C (RhoC). Oncotarget (2015) 6(7):6649–670. doi: 10.18632/oncotarget.6544

80. Shung S, Gong B, Wang P, Huang H, Luo L, Liu F. Novel Prognostic Biomarkers of Gastric Cancer Based on Gene Expression Microarray: COL1A2A, GSTA3, FGA and FGG. Mol Med Rep (2018) 18(4):3727–36. doi: 10.3892/mmr.2018.9368

81. Kamikawaji K, Seki N, Watanabe M, Matakai T, Takagi K, et al. Regulation of LOXL2 and SERPINH1 by Antitumor microRNA-29a in Lung Cancer With Idiopathic Pulmonary Fibrosis. J Cell Sci (2016) 6(3):68. doi: 10.1093/hmg/ddt365

82. Liu J, Huang BH, Zhang J, Carson DD, Hooi SC. Repression of HIP/PRL2 Expression Induces Differentiation in Colon Cancer Cells. J Cell Physiol (2006) 207(2):287–92. doi: 10.1002/jcp.20589

83. Chen D, Zhang R, Shen W, Fu H, Liu S, Sun K, et al. RPSL2-Specific shRNA Inhibits the Proliferation, Migration of BGC823 Gastric Cancer Cells With S100A4 as a Downstream Effector. Int J Oncol (2013) 42(5):1763–9. doi: 10.3892/ijo.2013.1872

84. Zhao X, Shen L, Feng Y, Yu H, Wu X, Chang J, et al. Decreased Expression of RPSL5A Suppresses Proliferation of Lung Cancer Cells. Tumour Biol (2015) 36(9):7673–40. doi: 10.1007/s13277-015-3577-9

85. Wang M, Hu Y, Stearns ME. RPSL2: A Novel Therapeutic Target in Prostate Cancer. J Exp Clin Cancer Res (2009) 286. doi: 10.1186/1755-7666-28-6

86. Dutton-Regester K, Gartner JJ, Emmanuel R, Quotob N, Davies MA, Gershwenfeld JE, et al. A Highly Recurrent RPS27 S’UTR Mutation in Melanoma. Oncotarget (2014) 5(10):2912–7. doi: 10.18632/oncotarget.2048
Cai J, Shi X, Li J, Tan F, Feng Y, Lin X, et al. Cystathionine β-Synthase Secretion Increases Peroxisome Proliferator-Activated Receptor γ Activity by Sulphydrylation at C139 Site Thereby Promoting Glucose Uptake and Lipid Storage in Adipocytes. Biochim Biophys Acta (2016) 1861(5):419–29. doi: 10.1016/j.bbalip.2016.03.001

Roudier E, Chapados N, Decary S, Gineste C, Le Bel C, Lavoie JM, et al. grill JI, Neumann J, Herbst A, Ofner A, Hiltwein F, Marschall MK, et al. Loss of function polymorphisms in the 5′-flanking region of the LEP gene: association with Leptin Deficiency Protects Against Non-Alcoholic Steatohepatitis. J Hepatol (2015) 63(1):68. doi: 10.1016/j.jhep.2014.09.014

Herder C, Hauner H, Kempf K, Kolb H, Skark T. Constitutive and Regulated Expression and Secretion of Interferon-γ-Inducible Protein 10 (IP-10)/CXCL10) in Human Adipocytes. Int J Obes (Lond) (2007) 31(3):403–10. doi: 10.1038/sj.ijo.0803432

Moua U, Onur H, Utkan G. Is Obesity Always a Risk Factor for All Breast Cancer Patients? c-erbB2 Expression is Significantly Lower in Obese Patients With Early Stage Breast Cancer. Clin Trans Oncol (2012) 14(12):923–30. doi: 10.1007/s12094-012-0878-z

Wang C, Xiang W, Xu P, Gu Y, Wang T, et al. Correlation of A2bAR and Mfge8 Promotes Obesity by Mediating the Uptake of Dietary Fats and Serum Fatty Acids. Cell Metab (2014) 20(2):175–83. doi: 10.1016/j.mce.2013.10.010

Joshi et al. Identification of Key Genes in Obesity

Frontiers in Endocrinology | www.frontiersin.org June 2021 | Volume 12 | Article 628907
