Midterm Part 1: Genome Wide Association Studies of Obesity

Brian Vail

Florida Atlantic University

Abstract

Obesity is a common disorder across the globe, involving excessive stores of body fat which leads to many health-related risks such as diabetes and heart disease. The Genome Wide Association Study has been the most recent leap in knowledge concerning obesity, identifying genes and gene products related to the disorder. In this article, previously identified genes related to obesity underwent GO term enrichment and biological pathway analysis. This resulted in the identification of biological pathway categories such as the regulation of biological quality, response to endogenous stimulus, and cytoplasmic vesicles.

Background

Obesity is described as a disorder involving an excessive amount of body fat. The negative effects on those individuals that suffer from obesity include heart disease, diabetes, high blood pressure and a host of other physical ailments (Mayo Clinic, 2019). Needless to say, this has become a major health issue around the globe. Many physicians, researchers, and scientists have studied obesity from every angle including treatment, recovery, and prevention. One of the most promising leads in the study of obesity is taking a genetic approach to one of the root causes of the disorder. The Genome Wide Association Study (GWAS) is an observational study where researchers study genetic variants across the genome to find correlations between genetic variants and traits. Generally, GWAS studies focus on human diseases such as obesity, studying any link between the disease and single-nucleotide polymorphisms (SNPs) in the genome (Wikipedia, 2019). In this chapter review by Hedman, Lindgren, and McCarthy (2014), they argue that despite an increased intake of nutrient-dense food and reduced physical activity, there is much evidence to support that obesity is also linked via familial genetics and a person's predisposition toward obesity. As such, they review and discuss how the GWAS approach has been used to link genetic variants and gene products to obesity and fat distribution. To analyze obesity, the authors focused on the phenotypic measurements of overall obesity (i.e. body mass index (BMI), fat percentage and extreme obesity) and fat distribution (waist-hip ratio (WHR), waist circumference (WC), and measures of visceral and cutaneous fat).

Methods

Literature Review

For the review section of this paper, the given exam chapter, “Genome Wide Association Studies of Obesity” was used to identify genes related to obesity and further discuss those relations. Pre-GWAS genes were identified from a previously given chapter, “Genetic Variation and Obesity Prior to the Era of Genome-Wide Association Studies”. Additional references were used, stemming from the assigned chapters.

Wikipedia and WebMD were utilized for further understanding of biological and medical concepts relevant to this review.

GO Term Enrichment

Identified genes from the literature review (both pre- and post-GWAS) were further identified on the HUGO Gene Nomenclature Committee (HGNC) website, confirming correct gene identification within their database. The Online Mendelian Inheritance in Man (OMIM) website was further used to identify genes. GO enrichment analysis was performed using Gene Ontology (www.geneontology.org) to determine related biological processes, cellular components, and molecular functions. To summarize and visualize the GO term results, GO terms and corresponding p values were entered into ReviGO to generate a scatterplot, interactive graph and treemap. The related GO terms were further explored using the GO database in Web Protégé and further concepts relating the GO terms to obesity were explored via Google Scholar.

The gene ontology database and Wikipedia were utilized for further understanding of relevant biological concepts.

Results

From the chapter review by McCormack (2014), a total of 25 pre-GWAS genes were identified as having links to obesity. These genes are summarized in Table 1.

Gene	HGNC ID	Approved Name	Gene Group
LEP	6553	leptin	
LEPR	6554	leptin receptor	
MC4R	6932	melanocortin 4 receptor	Melanocortin receptors
POMC	9201	proopiomelanocortin	Endogenous ligands
PCSK1	8743	proprotein convertase subtilisin/kexin type 1	Proprotein convertase subtilisin/kexin family
SNRPN	11164	small nuclear ribonucleoprotein polypeptide N	Sm spliceosomal proteins
NDN	7675	necdin	MAGE family
BBS1	966	Bardet-Biedl syndrome 1	Bardet-Biedl syndrome associated BBSome
AGRP	330	agouti related neuropeptide	Endogenous ligands
BDNF	1033	brain derived neurotrophic factor	Neurotrophins
SDC1	10658	syndecan 1	CD molecules, Syndecans
SDC3	10660	syndecan 3	Syndecans
SIM1	10882	SIM bHLH transcription factor 1	Basic helix-loop-helix proteins
CART	24323	CART prepropeptide	
UCP1	12517	uncoupling protein 1	Solute carriers

UCP3	12519	uncoupling protein 3	Solute carriers
GHRL	18129	ghrelin and obestatin prepropeptide	Endogenous ligands
GHSR	4267	growth hormone secretagogue receptor	Peptide receptors
PYY	9748	peptide YY	Endogenous ligands
PPARG	9236	peroxisome proliferator activated receptor gamma	Nuclear hormone receptors
PPARGC1B	30022	PPARG coactivator 1 beta	RNA binding motif containing, MicroRNA protein coding host genes
NR0B2	7961	nuclear receptor subfamily 0 group B member 2	Nuclear hormone receptors
ENPP1	3356	ectonucleotide pyrophosphatase/phosphodiesterase 1	Ectonucleotide pyrophosphatase/phosphodiesterase family
ADRB2	286	adrenoceptor beta 2	Adrenoceptors
ADRB3	288	adrenoceptor beta 3	Adrenoceptors

Table 1: Pre-GWAS Identified genes associated with obesity.

From the chapter review by Hedman, Lindgren, and McCarthy (2014), a total of 79 genes were discussed as having links to obesity. As the authors point out, some of the earliest links, pre-GWAS, have been found in the genes LEP, LEPR, and POMC which all took part in the hypothalamic circuitry of body weight regulation (Montague et al., 1997; Clement et al., 1998; Jackson et al., 1997). Specifically, LEP produces leptin and is involved in hormonal balance of energy balance mechanisms. LEPR provides a receptor protein for leptin. And POMC allows for a protein that is cleaved into multiple neuroendocrine messengers. Another early discovered gene is the melanocortin 4 receptor (MC4R) which is linked with severe, early-onset obesity (Farooqi

et al., 2000). However, while these early studies provided valued insights, the authors argue the greatest leap in the link between genetic variants and obesity has been seen in the GWAS.

Hedman, Lindgren, and McCarthy (2014) summarized the major findings that stemmed from GWAS that linked genetic variants to obesity. Specifically, they researched overall obesity including body mass index and fat percentage studies. As well, they investigated fat distribution studies, including studies of waist-hip ratio, waist circumference, and abdominal fat distribution. A summarized list of identified genes from these studies can be found in Table 2.

Gene	HGNC ID	Approved Name	Gene Group
ADAMTS9	13202	ADAM metallopeptidase with thrombospondin type 1 motif 9	ADAM metallopeptidases with thrombospondin type 1 motif
ADCY3	234	adenylate cyclase 3	Adenylate cyclases
ADCY9	240	adenylate cyclase 9	Adenylate cyclases
BDNF	1033	brain derived neurotrophic factor	Neurotrophins
CADM2	29849	cell adhesion molecule 2	C2, I, and V-set domain containing, Ig-like cell adhesion molecule family, MicroRNA protein coding host genes
CDKAL1	21050	CDK5 regulatory subunit associated protein 1 like 1	
CPEB4	21747	cytoplasmic polyadenylation element binding protein 4	RNA binding motif containing
DNM3	29125	dynamin 3	Pleckstrin homology domain containing MicroRNA protein coding host genes
ETV5	3494	ETS variant 5	ETS transcription factor family
FAIM2	17067	Fas apoptotic inhibitory molecule 2	Transmembrane BAX inhibitor motif containing

FANCL	20748	FA complementation group L	PHD finger proteins, FA complementation groups, FA Core Complex
FTO	24678	FTO alpha-ketoglutarate dependent dioxygenase	Alkylation repair homologs
GIPR	4271	gastric inhibitory polypeptide receptor	Glucagon receptor family, MicroRNA protein coding host genes
GNAT2	4394	G protein subunit alpha transducin 2	G protein subunits alpha, group i
GNPDA2	21526	glucosamine-6-phosphate deaminase 2	
GP2	4441	glycoprotein 2	
GPRC5B	13308	G protein-coupled receptor class C group 5 member B	G protein-coupled receptors, Class C orphans
GRB14	4565	growth factor receptor bound protein 14	Pleckstrin homology domain containing, SH2 domain containing
HNF4G	5026	hepatocyte nuclear factor 4 gamma	Nuclear hormone receptors
HOXB5	5116	homeobox B5	HOXL subclass homeoboxes
HOXC13	5125	homeobox C13	HOXL subclass homeoboxes
HS6ST3	19134	heparan sulfate 6-O-sulfotransferase 3	Sulfotransferases, membrane bound MicroRNA protein coding host genes
INSIG2	20452	insulin induced gene 2	
IRS1	6125	insulin receptor substrate 1	Pleckstrin homology domain containing
ITPR2	6181	inositol 1,4,5-trisphosphate receptor type 2	Inositol 1,4,5-triphosphate receptors
KCNMA1	6284	potassium calcium-activated channel subfamily M alpha 1	Potassium calcium-activated channels
KCTD15	23297	potassium channel tetramerization domain containing 15	KCTD family
KLF9	1123	Kruppel like factor 9	Kruppel like factors, Zinc fingers C2H2-type

LEP	6553	leptin	
LEPR	6554	leptin receptor	
LINGO2	21207	leucine rich repeat and Ig domain containing 2	I-set domain containing, MicroRNA protein coding host genes
LRP1B	6693	LDL receptor related protein 1B	Low density lipoprotein receptors
LY86	16837	lymphocyte antigen 86	
LYPLAL1	20440	lysophospholipase like 1	
MAF	6776	MAF bZIP transcription factor	Basic leucine zipper proteins
MAP2K5	6845	mitogen-activated protein kinase kinase 5	Mitogen-activated protein kinase kinases
MC4R	6932	melanocortin 4 receptor	Melanocortin receptors
MRPS33P4	29767	mitochondrial ribosomal protein S33 pseudogene 4	
MTCH2	17587	mitochondrial carrier 2	Solute carriers
MTIF3	29788	mitochondrial translational initiation factor 3	
NEGR1	17302	neuronal growth regulator 1	I-set domain containing IgLON cell adhesion molecules
NFE2L3	7783	nuclear factor, erythroid 2 like 3	Basic leucine zipper proteins
NISCH	18006	nischarin	
NPC1	7897	NPC intracellular cholesterol transporter 1	Solute carriers
NRXN3	8010	neurexin 3	Neurexins
NUDT3	8050	nudix hydrolase 3	Nudix hydrolase family
OLFM4	17190	olfactomedin 4	
PACS1	30032	phosphofurin acidic cluster sorting protein 1	
PCSK1	8743	proprotein convertase subtilisin/kexin type 1	Proprotein convertase subtilisin/kexin family
PIGC	8960	phosphatidylinositol glycan anchor biosynthesis class C	Phosphatidylinositol glycan anchor biosynthesis

POC5	26658	POC5 centriolar protein	
POMC	9201	proopiomelanocortin	Endogenous ligands
PRKCH	9403	protein kinase C eta	C2 domain containing protein kinases, AGC family kinases
PRKD1	9407	protein kinase D1	Pleckstrin homology domain containing
PTBP2	17662	polypyrimidine tract binding protein 2	RNA binding motif containing
PTER	9590	phosphotriesterase related	
RMST	29893	rhabdomyosarcoma 2 associated transcript	Long non-coding RNAs (published)
RPGRIP1L	29168	RPGRIP1 like	Protein phosphatase 1 regulatory subunits, C2 domain containing
RPL27A	10329	ribosomal protein L27a	L ribosomal proteins, snoRNA protein coding host genes
RPTOR	30287	regulatory associated protein of MTOR complex 1	MTOR complex 1, Armadillo-like helical domain containing, WD repeat domain containing
RSPO3	20866	R-spondin 3	Endogenous ligands
SDCCAG8	10671	serologically defined colon cancer antigen 8	Bardet-Biedl syndrome associated, MicroRNA protein coding host genes
SEC16B	30301	SEC16 homolog B, endoplasmic reticulum export factor	
SH2B1	30417	SH2B adaptor protein 1	Pleckstrin homology domain containing SH2 domain containing
SLC39A8	20862	solute carrier family 39 member 8	Solute carriers
SPRY2	11270	sprouty RTK signaling antagonist 2	
SSPN	11322	sarcospan	
STAB1	18628	stabilin 1	Scavenger receptors
TBX15	11594	T-box 15	T-boxes

TFAP2B	11743	transcription factor AP-2 beta	
TMEM160	26042	transmembrane protein 160	
TMEM18	25257	transmembrane protein 18	
TNKS	11941	tankyrase	Ankyrin repeat domain containing, Poly(ADP-ribose) polymerases, Sterile alpha motif domain containing, MicroRNA protein coding host genes
TNNI3K	19661	TNNI3 interacting kinase	
VEGFA	12680	vascular endothelial growth factor A	VEGF family
WARS2	12730	tryptophanyl tRNA synthetase 2, mitochondrial	Aminoacyl tRNA synthetases, Class I
ZNF608	29238	zinc finger protein 608	Zinc fingers C2H2-type
ZNRF3	18126	zinc and ring finger 3	Ring finger proteins
ZZZ3	24523	zinc finger ZZ-type containing 3	Zinc fingers ZZ-type, Myb/SANT domain containing, ATAC complex

Table 2: Identified genes associated with obesity.

GO term enrichment analysis resulted in 72 related biological pathway terms (27 biological processes; 45 cellular components). No significant results were found within the molecular function category. For ease of summary, only the overarching categories will be listed in the table below and the most prominent cases will be discussed.

The largest family of biological pathway terms comes under the regulation of biological quality heading. This is a process where the measured attributes of biological processes are controlled, such as size, shape and mass of parts of a cell. This is a very broad grouping and difficult relate a single instance of this to obesity without further refinement and biological knowledge. However,

this broader family included additional GO terms such as regulation of insulin secretion, regulation of bone remodeling, positive regulation of biological process and negative regulation of biological process. For example, in regulation of biological processes a biological process is either slowed, enhanced or completely stopped and is used as a means of bodily regulation. In a study by Yang et al. (2009), they explored regulation of metabolic processes in relation to obesity. They altered specific genes related to metabolic processes which contributed to abdominal obesity. Another interesting category in the biological processes is response to endogenous stimulus. This includes biological pathway terms such as response to cold, response to dietary excess and regulation of appetite. The body's regulation of appetite and its response to excessively large diet, not surprisingly, have a large impact on obesity.

The largest biological pathway category in the cellular components terms is cytoplasmic vesicles. This contained GO terms such as membrane-bounded organelle, platelet dense tubular network, which regulates platelet activation in blood platelets, and the Tle3-Aes complex, which is a transcriptional repressor complex. Another prominent family in the cellular components terms is the plasma membrane region. The postsynaptic membrane, as part of that family, is the part of neurons that accept messages in the form of neurotransmitters released from other neurons (cellular communication).

A summary of the identified pathways categories can be found in Table 3.

Type	Biological Pathway Category	GO Id	GO Description	Included GO Terms

Biological Process	behavior	GO:0007610	The internally coordinated responses (actions or inactions) of animals (individuals or groups) to internal or external stimuli, via a mechanism that involves nervous system activity. .	behavior
	feeding behavior	GO:0007631	Behavior associated with the intake of food.	feeding behavior adult feeding behavior
	response to endogenous stimulus	GO:0009719	Any process that results in a change in state or activity of a cell or an organism (in terms of movement, secretion, enzyme production, gene expression, etc.) as a result of a stimulus arising within the organism.	response to endogenous stimulus response to extracellular stimulus response to nitrogen compound response to oxygen-containing compound response to cold response to organic substance response to dietary excess regulation of appetite response to hormone regulation of response to food
	multicellular organismal process	GO:0032501	Any biological process, occurring at the level of a multicellular organism, pertinent to its function.	multicellular organismal process
	regulation of biological quality	GO:0065008	Any process that modulates a qualitative or quantitative trait of a biological quality. A biological quality is a measurable attribute of an organism or part of an organism, such as size, mass, shape, color, etc.	regulation of biological quality regulation of angiogenesis regulation of insulin secretion regulation of endothelial cell proliferation negative regulation of biological process positive regulation of biological process positive regulation of kinase activity glucose homeostasis regulation of hormone levels adaptive thermogenesis chemical homeostasis regulation of bone remodeling regulation of multicellular organismal process

Cellular Component	cellular_component	GO:0005575	A location, relative to cellular compartments and structures, occupied by a macromolecular machine when it carries out a molecular function. There are two ways in which the gene ontology describes locations of gene products: (1) relative to cellular structures (e.g., cytoplasmic side of plasma membrane) or compartments (e.g., mitochondrion), and (2) the stable macromolecular complexes of which they are parts (e.g., the ribosome).	cellular_component
	cell	GO:0005623	The basic structural and functional unit of all organisms. Includes the plasma membrane and any external encapsulating structures such as the cell wall and cell envelope.	cell
	membrane	GO:0016020	A lipid bilayer along with all the proteins and protein complexes embedded in it an attached to it.	membrane
	organelle	GO:0043226	Organized structure of distinctive morphology and function. Includes the nucleus, mitochondria, plastids, vacuoles, vesicles, ribosomes and the cytoskeleton, and prokaryotic structures such as anammoxosomes and pirellulosomes. Excludes the plasma membrane.	organelle
	receptor complex	GO:0043235	Any protein complex that undergoes combination with a hormone, neurotransmitter, drug or intracellular messenger to initiate a change in cell function.	receptor complex tubulobulbar complex TORC1 complex
	plasma membrane region	GO:0098590	A membrane that is a (regional) part of the plasma membrane.	plasma membrane region dendritic spine head postsynaptic endocytic zone membrane plasma membrane part postsynaptic endocytic zone insulin receptor complex plasma membrane integral component of plasma membrane postsynaptic membrane

	whole membrane	GO:0098805	Any lipid bilayer that completely encloses some structure, and all the proteins embedded in it or attached to it. Examples include the plasma membrane and most organelle membranes.	whole membrane
	membrane region	GO:0098589	A membrane that is a part of a larger membrane. Examples include the apical region of the plasma membrane of an epithelial cell and the various regions of the endoplasmic reticulum membrane.	membrane region integral component of vacuolar membrane intrinsic component of vacuolar membrane membrane raft
	photoreceptor inner segment	GO:0001917	The inner segment of a vertebrate photoreceptor containing mitochondria, ribosomes and membranes where opsin molecules are assembled and passed to be part of the outer segment discs.	photoreceptor inner segment
	endomembrane system	GO:0012505	A collection of membranous structures involved in transport within the cell. The main components of the endomembrane system are endoplasmic reticulum, Golgi bodies, vesicles, cell membrane and nuclear envelope.	endomembrane system cell part cell periphery cytoplasmic part cytoplasm
	neuron part	GO:0097458	Any constituent part of a neuron, the basic cellular unit of nervous tissue. A typical neuron consists of a cell body (often called the soma), an axon, and dendrites. Their purpose is to receive, conduct, and transmit impulses in the nervous system.	neuron part
	intrinsic component of membrane	GO:0031224	The component of a membrane consisting of the gene products having some covalently attached portion, for example part of a peptide sequence or some other covalently attached group such as a GPI anchor, which spans or is embedded in one or both leaflets of the membrane.	intrinsic component of membrane membrane part

	cytoplasmic vesicle	GO:0031410	A vesicle found in the cytoplasm of a cell.	cytoplasmic vesicle centrosome intracellular organelle part membrane-bounded organelle Tle3-Aes complex SREBP-SCAP-Insig complex intracellular membrane-bounded organelle organelle part glycosylphosphatidylinositol-N-acetylglucosaminyltransferase (GPI-GnT) complex axonemal microtubule BBSome platelet dense tubular network membrane vesicle lumen secretory granule lumen Golgi lumen

Table 3: Summarized enrichment results.

Discussion

Obesity is a widely studied disorder due to the serious implications on an individual's health. As the authors point out, over 500 million adults around the globe are obese (Hedman, Lindgren, and McCarthy, 2014; Finucane et al., 2011). One of the largest breakthrough in obesity research has come from the GWAS. And with advances in technology, identified genes can now be linked to related biological pathways via GO term enrichment. In this analysis, 25 pre-GWAS genes and 79 genes from the post-GWAS study that were related to obesity were identified and further analyzed using GO term enrichment to determine related biological pathways. This analysis results in 72 related biological pathways. Regulation of biological quality was one of the largest categories identified, and in particular negative regulation of biological processes. The negative

regulation of biological processes can have great impact on obesity. For example, negative regulation of white cell proliferation would have a large effect on obesity. White fat cells are used for energy storage, and when an individual takes in excess amounts of nutrition the white fat cells build up stored energy leading to obesity. When energy is needed, the fat cells are broken down through a series of chemical reactions that ultimately release fatty acids and metabolized by the body. Using enrichment analysis to pinpoint negative regulation of biological processes, such as white cell proliferation, could be of use to future researchers. This could be a starting point in terms of treating obesity in patients, such as by hypothetically increasing the negative regulation to decrease the amount of fat being stored. As well, this leads to further links, such as the fact that white fat cells are the location of the hormone leptin's production. This is of particular interest as the LEP gene, identified here as related to obesity, is responsible for the gene product leptin.

Another notable biological pathway category is glucose homeostasis. This refers to the maintenance of the internal state of glucose levels in the body. For example, glucose metabolism refers to the breaking down and usage of glucose, whether it is immediately used for energy or stored in the body as fat. This is a means of maintaining glucose homeostasis when too much glucose has been ingested, which is related to the identified GO term response dietary excess. While the link between glucose metabolism and obesity has long been established (Peiris et al., 1988), inspecting these relationships at a genetic level through enrichment analysis may allow for a new view to old issues. Such as considering the control of glucose homeostasis and metabolism through the regulatory proteins created at the transcription level. Future research could lead to the control of genetic factors of obesity that could be targeted directly to these

DNA transcriptional mechanisms to control which gene products are made and how much. Such as gene products that build up white fat cells from glucose could be decreased, or gene products that break down fat cells for energy could be increased.

Conclusions

Identifying genes and related biological pathways through GWAS and GO term enrichment allows a whole new resource for scientists, medical professionals and researchers to investigate obesity in terms of causes and potential treatment. Focusing on biological processes and pathways such as the regulation of biological quality, response to endogenous stimulus, and cytoplasmic vesicles identified through enrichment is crucial in furthering our understanding of obesity. Future work should continue to include previously unidentified genes and gene products and their associations with obesity.

References

Clement K., Vaisse C., Lahlou N., Cabrol S., Pelloux V., Cassuto D., . . . Guy-Grand B. (1998). A Mutation in Human Leptin Receptor Gene Causes Obesity and Pituitary Dysfunction. *Nature, 392(6674), 398-401*.

Farooqi I.S., Yeo G.S.H. Yeo, Keogh J.M., Aminian S., Jebb S.A., . . . O'Rahilly S. (2000). Dominant and Recessive Inheritance of Morbid Obesity Associated with Melanocortin 4 Receptor Deficiency. *Journal of Clinical Investigation, 106(2), 271-279*.

Finucane M.M., Stevens G.A., Cowan M.J., Danaei G., Lin J.K., Paciorek C.J., . . . Riley L.M. (2011). National, Regional, and Global Trends in Body-Mass Index Since 1980: Systematic Analysis of Health Examination Surveys and Epidemiological Studies With 960 Country-Years And 9·1 Million Participants. *Lancet, 377(9765), 557-567*.

Hedman A.K., Lindgren C.M., McCarthy M.I. (2014). Genome-Wide Association Studies of Obesity. *In The Genetics of Obesity*, Springer Science + Media, New York.

Jackson R.S., Creemers J.W., Ohagi S., Raffin-Sanson M.L., Sanders L., Montague C.T., . . . O'Rahilly S. (1997). Obesity and Impaired Prohormone Processing Associated with Mutations in the Human Prohormone Convertase 1 Gene. *Nature Genetics, 16(3), 303-306*.

Mayo Clinic. Obesity. Retrieved March 15th, 2019 from: https://www.mayoclinic.org/diseases-conditions/obesity/symptoms-causes/syc-20375742.

McCormack S.E. (2014). Genetic Variation and Obesity Prior to the Era of Genome-Wide Association Studies. *In The Genetics of Obesity*, Springer Science + Media, New York.

Montague C.T., Farooqi I.S., Whitehead J.P., Soos M.A., Rau H., Wareham N.J., . . . O'Rahilly S. (1997). Congenital Leptin Deficiency is Associated with Severe Early-Onset Obesity in Humans. *Nature, 387(6636), 903-908*.

Peiris A.N., Struve M.F., Mueller R.A., Lee M.B., Kissebah A.H. (1988). Glucose Metabolism in Obesity: Influence of Body Fat Distribution. *The Journal of Clinical Endocrinology & Metabolism, 67(4), 760-767*.

Wikipedia. Genome-wide association study. Retrieved March 15th, 2019 from: https://en.wikipedia.org/wiki/Genome-wide_association_study.

Yang X., Deignan J.L., Qi H., Zhu J., Qian S., Zhong, J., . . . Drake T.A. (2009). Validation of Candidate Casual Genes for Abdominal Obesity Which Affect Shared Metabolic Pathways and Networks. *Nature Genetics, 41(4), 415-423*.

Midterm Part 2: Genome Wide Association Studies of Obesity and the Related Genes and Biological Pathways to the Preventive Treatment of Hypercholesterolemia and Atherosclerosis

Brian Vail

Florida Atlantic University

Abstract

Comparing the gene enrichment analyses conducted on obesity and atherosclerosis shows few commonalities in biological pathways that contribute to these disorders. Lipid and cholesterol metabolism are related to chemical homeostasis and, specifically low-density lipoprotein levels, may play large roles in both obesity and atherosclerosis. As well, the LEP gene, which was not previously identified with atherosclerosis, has been found to not only associate with obesity but to also play a large role in atherosclerosis. This comparative analysis further discusses the similarities and shared root genetic causes of obesity and atherosclerosis.

Background

Two of the largest risk factors of atherosclerosis and coronary heart disease include hypercholesterolemia and obesity (Rooy and Pretorius, 2014). Emanuela et al. (2012) suggest that obesity leads to hypertension and hypercholesterolemia which eventually leads to atherosclerosis. Other studies have found that obesity and adipose tissues (i.e. white fat cells) cause a pro-inflammatory environment which ultimately lead to atherosclerosis (Lyon, Law and Hsueh, 2003). As well, with hypertension being associated with obesity, others have found links between hypertension and increased cholesterol quantities. It was found that patients were likely to develop hypertension when they had increased levels of low-density lipoproteins (LDL) and decreased levels of high-density lipoproteins (HDL). In these cases, increased cholesterol dramatically increased the risk of coronary heart disease (Borghi, Grassi and Taddei, 2008). This is important as previous research into atherosclerosis showed several gene variants (LPA, APOB, SORT1, LDLR, APOE, ABCG5 – ABCG8, and PCSK9) identified in association with atherosclerosis and hypercholesterolemia were also associated with high levels of low-density lipoprotein cholesterol (LDL-C) which lead to increased risk of atherosclerotic cardiovascular disease (Roberts, 2014).

Looking a different view of obesity and atherosclerosis, Turkbey et al. (2010) sought to investigate the impact of obesity on the left ventricle of the heart based on participants diagnosed with atherosclerosis. They based this on the fact that in the Framingham studies, it was found that obesity was related to an increase in wall thickness of the heart, more so than an increase in chamber size (Lauer et al., 1991). This is likely because the heart, as a muscle, must work harder

to pump blood through obese patients. The researchers in fact found that from the members of the atherosclerosis group obesity was associated with increased ventricular thickness.

Taking into account the connections between obesity, LDL-C and genetic variants associated with atherosclerosis, it is expected that a large degree of overlap would be seen between the root genetic causes and biological pathways that contribute to obesity, atherosclerosis and hypercholesterolemia. Taking into account the additional perspective of obesity at the genetic level, this could lead to new insights in the treatment and prevention of atherosclerosis and hypercholesterolemia.

Methods

Literature Review

For the literature review section of this paper, research articles were found using the online resources Google Scholar, The National Center for Biotechnology Information (NCBI), and PubMed. Key terms used for search results included “Genetics”, “Atherosclerosis”, “Hypercholesteremia”, “Prevention”, “Treatment”, “Risk Factors”, “Obesity”, and others.

Wikipedia and WebMD were utilized for further understanding of biological and medical concepts relevant to this review.

GO Term Enrichment

Identified genes from the literature reviews and the previous assignment were cross-referenced to identify any duplicates across obesity studies and preventative treatment of hypercholesterolemia

and atherosclerosis. Based on GO enrichment analyses performed for part 1 of the midterm and assignment 2, the lists of GO terms identified via Gene Ontology (www.geneontology.org) were combined into a single Microsoft Excel file. Base on GO identification numbers, a duplicates function was used to identify GO terms that appeared in both enrichment analyses. The list of combined GO terms was not summarized using ReviGO because 1) The list of duplicate GO terms was not large and 2) The duplicate GO terms had different p values depending on their source (Obesity vs. Atherosclerosis).

Identified genes from the literature reviews (pre- and post-GWAS) and the previous assignment were combined into a single list and underwent GO enrichment analysis using Gene Ontology (www.geneontology.org) to determine related biological processes, cellular components, and molecular functions. To summarize and visualize the GO term results, GO terms and corresponding p values were entered into ReviGO to generate a scatterplot, interactive graph and treemap. The related GO terms were further explored using the GO database in Web Protégé and further concepts relating the GO terms to obesity were explored via Google Scholar.

Results

When comparing the gene identification parts of the obesity review article and the previous hypercholesterolemia and atherosclerosis research, no common genes were found. When comparing obesity with hypercholesterolemia and atherosclerosis enrichment GO terms, a select few biological pathways were common in both. However, the terms which included "endomembrane system", "cytoplasmic vesicle", "cytoplasmic vesicle part", "vacuolar part",

“chemical homeostasis”, and “intracellular vesicle” are very broad terms which do not allow for specific connections between obesity and atherosclerosis. For this reason, speculation about possible links between obesity and atherosclerosis will be the focus. The combined GO enrichment analysis of the original gene lists resulted in 167 related biological pathway terms (138 biological processes; 14 cellular components; 15 molecular functions). For ease of summary, only the overarching categories will be listed in the table below and the most prominent cases will be discussed.

Chemical homeostasis, as found to be a common biological pathway between the obesity and atherosclerosis studies, refers to the biological processes involved in maintaining an internal balance of chemicals in the body. Related to that, one could argue the cholesterol metabolic process is a result of chemical homeostasis. This would make sense, as increased levels of LDL-C, the very nature of hypercholesterolemia, are associated with both obesity and atherosclerosis. Also, this would agree with the results of the combined gene list enrichment. Biological pathways including “sterol and steroid metabolic process”, “regulation of lipid metabolic process”, “lipid homeostasis”, “cholesterol binding”, and others were identified which lends evidence that the maintenance of LDL-C levels if associated with these disorders. This link further solidifies the idea that obesity and atherosclerosis are related. In a study by Ouimet (2013), investigated the process of autophagy, where the body removes damaged cells and regenerates new ones, in relation to obesity and atherosclerosis. It is pointed out that lipid metabolism is at the heart of these disorders and that autophagy and lipid homeostasis are important in the role of lipid regulation. Ouimet (2013) proposes that manipulation of autophagic

activity could lead to new forms of treatment and prevention of disorders such as obesity and atherosclerosis by helping to regulate lipid metabolism.

Other GO terms of note that were identified in the combined gene list enrichment include "response to dietary excess", "regulation of appetite", and "feeding behavior". The terms all suggest the obvious, that diet has an impact of obesity and, ultimately, atherosclerosis. And while the obvious suggestion of diet control as a means of preventative care, future work can focus on the biological pathways in which food is metabolized and either used or stored as a means of reducing obesity and atherosclerosis.

A summary of joint identified pathways can be found in Table 1 and the combined gene list enrichment results in Table 2.

Biological Pathway	**GO Id**	**GO Description**
endomembrane system	GO:0012505	A collection of membranous structures involved in transport within the cell. The main components of the endomembrane system are endoplasmic reticulum, Golgi bodies, vesicles, cell membrane and nuclear envelope. Members of the endomembrane system pass materials through each other or though the use of vesicles.
cytoplasmic vesicle	GO:0031410	A vesicle found in the cytoplasm of a cell.
cytoplasmic vesicle part	GO:0044433	Any constituent part of cytoplasmic vesicle, a vesicle formed of membrane or protein, found in the cytoplasm of a cell.
vacuolar part	GO:0044437	Any constituent part of a vacuole, a closed structure, found only in eukaryotic cells, that is completely surrounded by unit membrane and contains liquid material.
chemical homeostasis	GO:0048878	Any biological process involved in the maintenance of an internal steady state of a chemical.

intracellular vesicle	GO:0097708	Any vesicle that is part of the intracellular region.

Table 1: Combined GO pathways.

Type	Biological Pathway Category	GO Id	GO Description	Included GO Terms
Biological Process	behavior	GO:0007610	The internally coordinated responses (actions or inactions) of animals (individuals or groups) to internal or external stimuli, via a mechanism that involves nervous system activity.	behavior
	biological_process	GO:0008150	A biological process represents a specific objective that the organism is genetically programmed to achieve.	biological_process
	cellular process	GO:0009987	Any process that is carried out at the cellular level, but not necessarily restricted to a single cell.	cellular process
	response to extracellular stimulus	GO:0009991	Any process that results in a change in state or activity of a cell or an organism (in terms of movement, secretion, enzyme production, gene expression, etc.) as a result of an extracellular stimulus.	response to extracellular stimulus cellular response to oxygen-containing compound response to nitrogen compound response to oxygen-containing compound cellular response to stimulus response to hormone response to endogenous stimulus response to dietary excess regulation of appetite response to chemical response to organic substance

	sterol metabolic process	GO:0016125	The chemical reactions and pathways involving sterols, steroids with one or more hydroxyl groups and a hydrocarbon side-chain in the molecule.	sterol metabolic process secondary alcohol metabolic process steroid metabolic process regulation of lipid metabolic process neutral lipid metabolic process triglyceride metabolic process lipid metabolic process diterpenoid metabolic process retinoid metabolic process glycerolipid catabolic process
	signaling	GO:0023052	The entirety of a process in which information is transmitted within a biological system.	signaling
	multicellular organismal process	GO:0032501	Any biological process, occurring at the level of a multicellular organism, pertinent to its function.	multicellular organismal process
	developmental process	GO:0032502	A biological process whose specific outcome is the progression of an integrated living unit: an anatomical structure (which may be a subcellular structure, cell, tissue, or organ), or organism over time from an initial condition to a later condition.	developmental process
	macromolecular complex remodeling	GO:0034367	The acquisition, loss, or modification of macromolecules within a complex, resulting in the alteration of an existing complex.	macromolecular complex remodeling feeding behavior protein-lipid complex subunit organization plasma lipoprotein particle remodeling regulation of multicellular organismal process regulation of angiogenesis regulation of digestive system process adaptive thermogenesis adult feeding behavior

	response to stimulus	GO:0050896	Any process that results in a change in state or activity of a cell or an organism (in terms of movement, secretion, enzyme production, gene expression, etc.) as a result of a stimulus.	response to stimulus
	biological regulation	GO:0065007	Any process that modulates a measurable attribute of any biological process, quality or function.	biological regulation
	regulation of biological quality	GO:0065008	Any process that modulates a qualitative or quantitative trait of a biological quality. A biological quality is a measurable attribute of an organism or part of an organism, such as size, mass, shape, color, etc.	regulation of biological quality regulation of endothelial cell proliferation lipid homeostasis negative regulation of cellular process positive regulation of cellular process negative regulation of biological process positive regulation of biological process regulation of signaling positive regulation of cellular catabolic process positive regulation of lipid metabolic process regulation of lipase activity regulation of cell communication regulation of lipoprotein lipase activity regulation of localization chemical homeostasis regulation of developmental process regulation of insulin secretion regulation of catalytic activity regulation of biological process regulation of plasma lipoprotein particle levels regulation of hormone levels cholesterol homeostasis activation of adenylate cyclase activity

	biosynthetic process	GO:0009058	The chemical reactions and pathways resulting in the formation of substances; typically the energy-requiring part of metabolism in which simpler substances are transformed into more complex ones.	biosynthetic process
	cell communication	GO:0007154	Any process that mediates interactions between a cell and its surroundings.	cell communication
	lipoprotein metabolic process	GO:0042157	The chemical reactions and pathways involving any conjugated, water-soluble protein in which the covalently attached nonprotein group consists of a lipid or lipids.	lipoprotein metabolic process
	organic hydroxy compound metabolic process	GO:1901615	The chemical reactions and pathways involving organic hydroxy compound.	organic hydroxy compound metabolic process
	cholesterol efflux	GO:0033344	The directed movement of cholesterol, cholest-5-en-3-beta-ol, out of a cell or organelle.	cholesterol efflux organic hydroxy compound transport lipid localization regulation of lipid transport phospholipid efflux
	lipoprotein catabolic process	GO:0042159	The chemical reactions and pathways resulting in the breakdown of any conjugated, water-soluble protein in which the covalently attached nonprotein group consists of a lipid or lipids.	lipoprotein catabolic process
	organic substance biosynthetic process	GO:1901576	The chemical reactions and pathways resulting in the formation of an organic substance, any molecular entity containing carbon.	organic substance biosynthetic process

Cellular Component	endomembrane system	GO:0012505	A collection of membranous structures involved in transport within the cell.	endomembrane system
	plasma lipoprotein particle	GO:0034358	A spherical particle with a hydrophobic core of triglycerides and/or cholesterol esters, surrounded by an amphipathic monolayer of phospholipids, cholesterol and apolipoproteins.	plasma lipoprotein particle protein-lipid complex
	cytoplasmic vesicle	GO:0031410	A vesicle found in the cytoplasm of a cell.	cytoplasmic vesicle membrane-bounded organelle
Molecular Function	Signal receptor binding	GO:0005102	Interacting selectively and non-covalently with one or more specific sites on a receptor molecule, a macromolecule that undergoes combination with a hormone, neurotransmitter, drug or intracellular messenger to initiate a change in cell function.	receptor binding lipoprotein particle receptor binding G-protein coupled receptor binding low-density lipoprotein particle receptor binding
	binding	GO:0005488	The selective, non-covalent, often stoichiometric, interaction of a molecule with one or more specific sites on another molecule.	binding
	sterol transporter activity	GO:0015248	Enables the directed movement of sterols into, out of or within a cell, or between cells. Sterol are steroids with one or more hydroxyl groups and a hydrocarbon side-chain in the molecule.	sterol transporter activity cholesterol transporter activity

	phosphatidylcholine-sterol O-acyltransferase activator activity	GO:0060228	Increases the activity of phosphatidylcholine-sterol O-acyltransferase, an enzyme that converts cholesterol and phosphatidylcholine (lecithins) to cholesteryl esters and lyso-phosphatidylcholines.	phosphatidylcholine-sterol O-acyltransferase activator activity
	lipoprotein particle binding	GO:0071813	Interacting selectively and non-covalently with a lipoprotein particle.	lipoprotein particle binding protein-lipid complex binding
	alcohol binding	GO:0043178	Interacting selectively and non-covalently with an alcohol, any of a class of alkyl compounds containing a hydroxyl group.	alcohol binding cholesterol binding peptide binding protein binding

Table 2: Combined GO enrichment analysis.

Discussion

Both obesity and atherosclerosis have been linked with high cholesterol, specifically increased levels of LDL-C and decreased levels of HDL-C. One of the root causes of atherosclerosis is long-term, heightened levels of LDLs where the cholesterol builds up in the arteries and blocks blood flow as plaques. Likewise, high levels of LDLs are found in obese patients (Borghi, Grassi and Taddei, 2008). However, the key difference is that atherosclerosis is caused by increased levels of LDLs, while obesity leads to increased levels of LDLs. As part of chemical homeostasis, and with further study, the enrichment analysis could lead the way to confirm what is already well known and accepted: obesity is a risk factor for hypercholesterolemia and atherosclerosis. However, when concerned with preventative treatment of hypercholesterolemia and atherosclerosis at the genetic level, this opens up more treatment possibilities and warrants

continued research. Statins are historically the most widely used form of treatment in lowering cholesterol levels and decreasing the risk of hypercholesterolemia and atherosclerosis. A recent study confirmed this when they found that statins lowered the risk of a cardiovascular event in those at high-risk of coronary heart disease (Thavendiranathan, Bagai, Brookhart, & Choudhry, 2006). However, when investigating the causes of obesity, further research should be conducted in preventative treatment. Advances here could be used to decrease obesity rates, thus helping to decrease elevated LDL level and decreasing the risk of atherosclerosis. Furthermore, a major component to obesity has to do with cholesterol metabolism and catabolism. The rate at which cholesterol is broken down, stored for energy, and depleted could, in theory, be a vulnerable spot when it comes to manipulation of biological processes as a means of preventative treatment.

The LEP gene was identified in the obesity study, however it was not yet identified in the hypercholesterolemia and atherosclerosis study. It's been found that adipose tissue can, in part, lead to atherosclerosis by way of secreting cytokines like leptin (Lyon et al., 2003). Additionally, leptin has been shown to increasing the macrophage process of cholesterol uptake (Lyon et al., 2003; O'Rourke et al., 2002). This suggests that while the LEP gene was already associated with obesity, it should also be associated with atherosclerosis and hypercholesterolemia.

Other commonalities were identified, but the terms such as "cytoplasmic vesicle", "endomembrane system", and "vacuolar part" are broad terms that apply to many different biological pathways. While a literature review did not uncover anything directly related to these generalized terms, they are important to keep in mind when closely inspecting other, more distinct, identified biological pathways. Further investigation of the combined gene list enrichment results is warranted by experts in the field is warranted as there were over 150

biological pathways identified, many of which were much more specific than the results of the joint pathways. Specifically, “sterol metabolic processes”, and “macromolecular complex remodeling” including GO terms related to lipoproteins, sterols, and cholesterol warrant further study in relating obesity and atherosclerosis.

Conclusions

While several genes and biological pathways were previously identified in hypercholesterolemia and atherosclerosis, investigating the genetic basis for obesity led to further insights to atherosclerosis. Advances which may potentially lead to the discovery of new forms of preventative treatment by targeting previously undiscovered genes and biological pathways. One such potential lies in reducing LDL-C levels in a patient by preventing obesity, either by biological manipulation or lifestyle changes. Thus, targeting some of the biological pathways that contribute to obesity may provide alternative forms of preventative treatment for atherosclerosis. Another such potential lies in further exploring the LEP gene that was not previously identified as being associated with atherosclerosis. Future work should focus on the related aspects of obesity with the aim in preventing obesity as a means of preventing hypercholesterolemia and atherosclerosis.

References

Borghi C., Grassi G., Taddei S. (2008). Interactions Between Hypertension and Hypercholesterolemia: From Epidemiology to Therapeutic Implications. *Current Hypertension Reviews, 4(1), 1-7*.

Emanuela F., Grazia M., Marco D.R., Paoloa L.M., Giorgio F., Marco B. (2012). Inflammation as a Link Between Obesity and Metabolic Syndrome. *Journal of Nutrition and Metabolism, 2012*.

Lauer M.S., Anderson K.M., Kannel W.B., Levy D. (1991). The Impact of Obesity on Left Ventricular Mass and Geometry. *The Journal of the American Medical Association, 266, 231-236*.

Lyon C., Law R.E., Hsueh W.A. (2003). Minireview: Adiposity, Inflammation and Atherogenesis. *Endocrinal, 144(6), 2195-2200*.

O'Rourke L., Gronning L.M., Yeaman S.J., Shepherd P.R. (2002). Glucose-Dependent Regulation of Cholesterol Ester Metabolism in Macrophages by Insulin and Leptin. Journal of Biological Chemistry, 277(45), 42557-42562.

Ouimet M. (2013). Autophagy in Obesity and Atherosclerosis: Interrelationships between Cholesterol Homeostasis, Lipoprotein Metabolism and Autophagy in Macrophages and Other Systems. *BBA – Molecular and Cell Biology of Lipids, 1831(6), 1124-1133*.

Roberts, R. (2014). Genetic of Coronary Artery Disease: An Update. *Methodist Debakey Cardiovascular Journal, 10(1), 7-12*.

Rooy M.J., Pretorius E. (2014). Obesity, Hypertension and Hypercholesterolemia as Risk Factors for Atherosclerosis Leading to Ischemic Events. *Current Medical Chemistry, 21, 2121-2129*.

Thavendiranathan P., Bagai A., Brookhart M.A., & Choudhry N.K. (2006). Primary Prevention of Cardiovascular Diseases with Statin Therapy: A Meta-analysis of Randomized Controlled Trials. *Archive of Internal Medicine*, 166(21), 2307–2313.

Turkbey E.B., McClelland R.L., Kronmal R.A., Burke G.L., Bild D.E., Tracy R.P., . . . Bluemke D.A. (2010). The Imapct of Obesity on the Left Ventricle. The Multi-Ethnic Study of Atherosclerosis (MESA). JACC: *Cardiovascular Imaging, 3(3)*.