

RESEARCH ARTICLE

# Comprehensive Map of Molecules Implicated in Obesity

Jaisri Jagannadham<sup>1a</sup>, Hitesh Kumar Jaiswal<sup>2</sup>, Stuti Agrawal<sup>2b</sup>, Kamal Rawal<sup>1\*</sup>

Department of Biotechnology, Jaypee Institute of Information Technology, Noida [UP]-201 307, India

☞ These authors contributed equally to this work.

<sup>1a</sup> Current address: Jaipur National University, Jaipur, India

<sup>2b</sup> Current address: Center for Data Intensive Science, University of Chicago, Chicago, Illinois, United States of America

\* [kamal.rawal@gmail.com](mailto:kamal.rawal@gmail.com)

## Abstract

Obesity is a global epidemic affecting over 1.5 billion people and is one of the risk factors for several diseases such as type 2 diabetes mellitus and hypertension. We have constructed a comprehensive map of the molecules reported to be implicated in obesity. A deep curation strategy was complemented by a novel semi-automated text mining system in order to screen 1,000 full-length research articles and over 90,000 abstracts that are relevant to obesity. We obtain a scale free network of 804 nodes and 971 edges, composed of 510 proteins, 115 genes, 62 complexes, 23 RNA molecules, 83 simple molecules, 3 phenotype and 3 drugs in “bow-tie” architecture. We classify this network into 5 modules and identify new links between the recently discovered fat mass and obesity associated FTO gene with well studied examples such as insulin and leptin. We further built an automated docking pipeline to dock orlistat as well as other drugs against the 24,000 proteins in the human structural proteome to explain the therapeutics and side effects at a network level. Based upon our experiments, we propose that therapeutic effect comes through the binding of one drug with several molecules in target network, and the binding propensity is both statistically significant and different in comparison with any other part of human structural proteome.



CrossMark  
click for updates

## OPEN ACCESS

**Citation:** Jagannadham J, Jaiswal HK, Agrawal S, Rawal K (2016) Comprehensive Map of Molecules Implicated in Obesity. PLoS ONE 11(2): e0146759. doi:10.1371/journal.pone.0146759

**Editor:** Filippo Castiglione, National Research Council of Italy (CNR), ITALY

**Received:** August 26, 2014

**Accepted:** December 22, 2015

**Published:** February 17, 2016

**Copyright:** © 2016 Jagannadham et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** The authors have no support or funding to report.

**Competing Interests:** The authors have declared that no competing interests exist.

## Introduction

Obesity, a complex condition with serious medical, psychological and social consequences, affects millions of people across the world [1]. In addition, rising numbers of juvenile onset obesity cases contribute to increased incidence of time-dependent complications of obesity such as insulin resistance, non-insulin-dependent diabetes mellitus, hypertension, coronary artery disease and other cardiac disorders often grouped as “metabolic syndrome X” [2–3]. The pathophysiology of obesity is influenced by several factors such as candidate genes and their expression, single nucleotide polymorphisms, proteins, metabolic pathways and their perturbations due to mutations, nutrition, exercise, gut microbes, and diseases, e.g. hypothyroidism [4–5]. Experts recommend that increase in physical activity and reduction in intake of high calorie

foods, can act as possible deterrent to obesity epidemic. Numerous studies have examined the use of medicines [6–7] and surgery [8] as possible treatment measures, although clinical studies also indicate that recurrences are high in people who have lost weight through diet, exercise or medication [9].

Adipose tissue is central to the regulation of energy balance. In the pathophysiology of obesity, chronic adipose tissue inflammation is a hallmark [10]. Two functionally different types of fats are present in mammals: white and brown adipose tissue. White adipose tissue is the primary site of triglyceride storage whereas brown adipose tissue is implicated in energy expenditure. The latter has an ability to counteract obesity [11]. Adipogenesis, namely the formation of adipose tissue begins with the commitment of mesenchymal stem cells (MSCs) to the adipocyte lineage, followed by terminal differentiation of preadipocytes to mature adipocytes [12]. The adipose tissue storage is influenced by environmental and genetic factors. The environmental influence generally depends on the individual's life-style, for instance, food intake and physical activity. The importance of genetic factors in obesity has been demonstrated by twin studies, adoption studies and segregation analyses [13–14].

The literature data pertaining to obesity is vast and complex. For instance, 'obesity' as a keyword in PubMed yielded over 13,000 results (hits) for a single year 2011. It was increased by 11,612 hits in 2012 and 11,177 hits in 2013, showing that literature data is growing at rapid pace. In addition to proteins and other molecules, these abstracts contain reports from clinical, genetic, mutational and meta-studies. To construct comprehensive molecular map of molecules reported in obesity, we used information from full length articles using deep curation model [15]. A deep curation model performs better than text mining methods, particularly in terms of accuracy, but has the disadvantage of being labour intensive and time consuming [16]. Given this constraint, it is difficult to curate large number of papers published every year and such a resource will become obsolete in absence of regular revision and updates by experts. Therefore, we decided to develop a hybrid system combining text mining systems and deep curation strategy to screen large amount of published data available on obesity to provide up-to-date information.

Networks pervade our lives as exemplified by worldwide webs, internet, small world networks, electricity grids, social networks, topology of food webs, citation networks as well as metabolic networks. To understand the role of networks in complex diseases, there were several attempts to construct disease networks [17–19]. A research group built a pathway on an autoimmune disease, "Rheumatoid arthritis" using microarray data [20]. In biological systems, at molecular or cellular levels, several reconstructions of comprehensive pathways have been conducted using published literature data. These include compiling human cell-cycle events by Kohn [21], comprehensive maps of EGFR pathway [22], Toll-like receptor signalling pathway [23] and RB/E2F pathway [24]. Apart from these, researchers have also used microarray data [25], protein-protein interaction data [26–27], co-cited data [28] as well as literature data [29] to construct networks. Despite all these efforts, there is plenty of scope to expand the role of networks in disease pathophysiology.

Human Obesity Gene Map 2005 is considered to be one of the best information resource for genes implicated in obesity [30]. The Human Obesity Gene Map 2005 provides evidence from single-gene mutation obesity cases, Mendelian disorders exhibiting obesity as a clinical feature, transgenic and knockout murine models relevant to obesity, quantitative trait loci (QTL) from animal cross-breeding experiments, association studies with candidate genes, and linkages from genome scans and genes or markers that have been shown to be associated or linked with obesity phenotype. We identified 379 genes reported in obesity from Human Obesity Gene Map and included in our proposed network. Transcription factors play an important role in conversion of pre-adipocytes to adipocytes and involved in several other mechanisms

pertaining to obesity pathophysiology. For this reason, we retrieved 114 transcription factors from DGAP (Diabetes Genome Anatomy Project) & GenMapp (<http://www.genmapp.org/default.html>). Apart from these, we identified an independent set of 33 genes reported in obesity literature (See Table A in [S2 File](#) for list of molecules & Table B in [S2 File](#) for experimental evidence). This work was complemented by mining over 35,000 genes in 96,219 abstracts using perl scripts. Through text-mining, we found 4,274 genes as first round of 'hits' (See [Methods](#)). Since text mining systems are known to produce large number of false hits, therefore we screened these hits manually and removed gene names matching with common English words, abbreviations and methodology terms using various types of filters (See examples provided on our website <http://tinyurl.com/d74r9xy> as well as in [S3 File](#)). Out of 4,274 hits, we label 1,268 genes as positive hits and 3,006 as false positive hits (See Table C in [S2 File](#)). Text mining system also reported several recently published molecules such as fat mass and obesity associated (FTO) and omega-3 fatty acid receptor 1 (GPR120) [[31–32](#)].

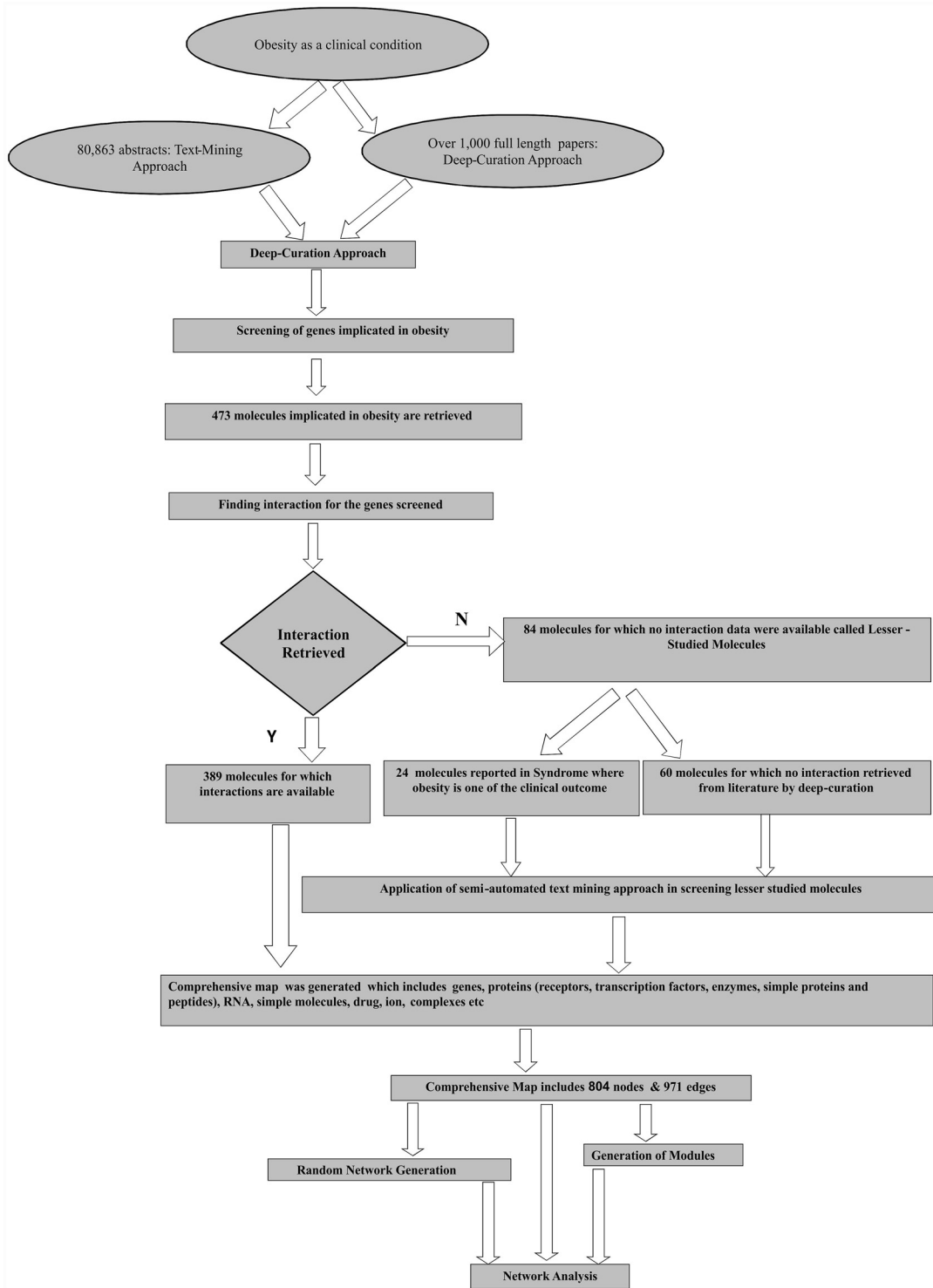
Based upon these techniques, we constructed two datasets (A and B) to create comprehensive network. Set A consist of 473 genes and proteins retrieved through deep curation strategy whereas set B consist of 1268 genes retrieved through semi-automatic text mining system. We started with this set of molecules as a 'partial list' of the proposed comprehensive network and expanded by adding more molecules based upon interactions reported in literature in context of obesity. The final comprehensive map was constructed based upon genes, proteins, receptors, transcription factors, enzymes, ion channels, drugs, RNA molecules, simple molecules and their relationships (See [Fig 1](#)).

## Results

### General features of the map

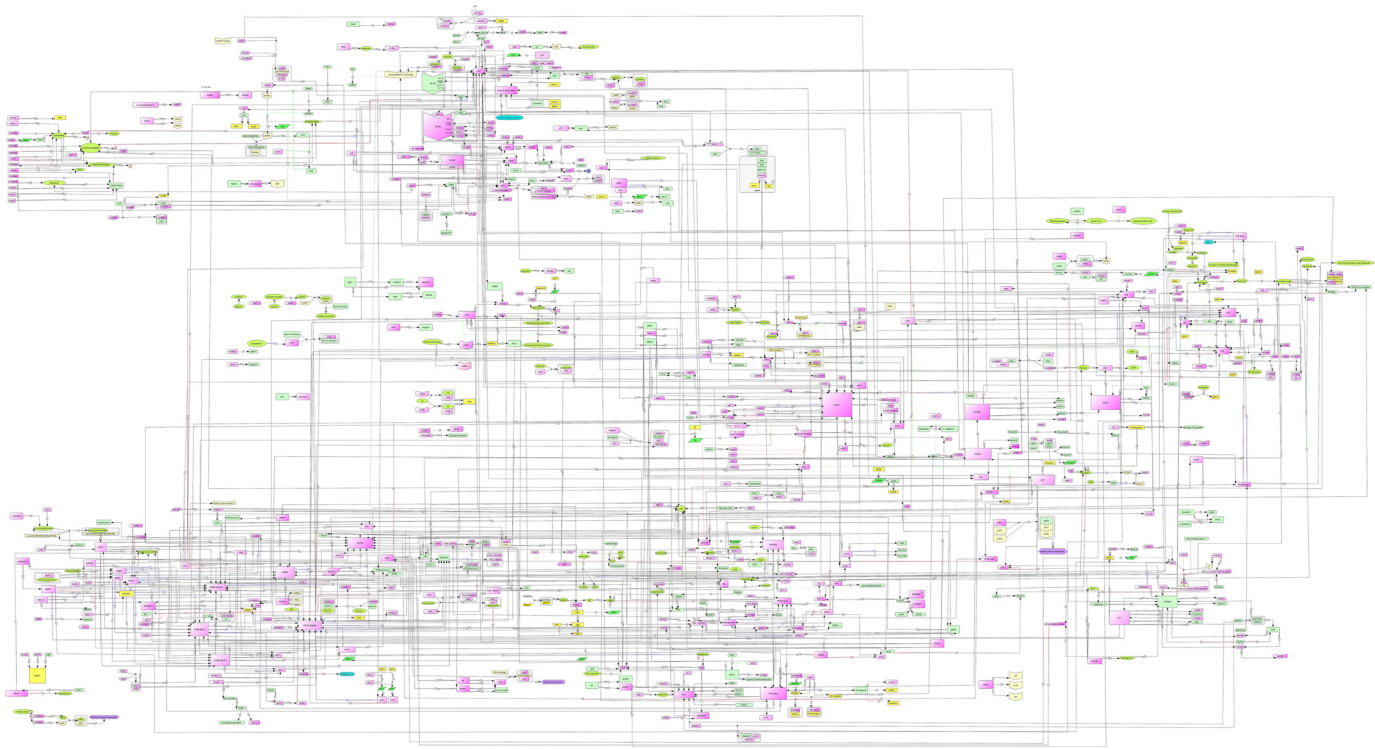
We screened over one thousand research articles manually and more than 96,219 abstracts (published till December 2012) using text mining system. The majority of molecules identified in this study can be tracked to sources such as human obesity gene map database update 2005, GenMapp and miscellaneous literature reports (Table A in [S2 File](#)). We have prepared a resource base where each molecule is linked with its research article. Each paper is curated manually and the portion of text denoting gene (molecule) or its interaction with other molecules in context of obesity is highlighted. The information on interactions of molecule is given on our website (Supplementary Folder 1: Gene interaction evidence: "<http://tinyurl.com/nc3yjj7>" & A in [S3 File](#) & Table D in [S2 File](#)). During this study, we encountered set of molecules which are found to be involved in syndromes where obesity is one of the clinical outcomes. Since the direct evidence on the role of these molecules in obesity is not known, we decided to include them as an independent part of the proposed map. We label this set as 'lesser studied (reported) group' due to paucity of literature data. To illustrate, ALMS1 gene is related with an "Alstrom syndrome 1" where obesity is a frequent clinical outcome in patients [[33](#)], but, direct evidence of linking ALMS1 with obesity is not reported in literature. Similarly, Gamma-aminobutyric acid A receptor gamma 3 (GABRG3) is an early childhood obesity gene reported in Prader-Willi syndrome [[34](#)], but, direct experimental evidence is not known. Likewise, genes reported from X-linked mutation studies and linkage studies could not be placed in the main network due to sparse experimental or interaction data. Therefore, out of 473 molecules, we included 389 molecules in the proposed network and the rest 84 molecules were reported as an independent set (See Table E in [S2 File](#)). The process, of incorporation of lesser studied molecules in proposed network, is elaborated in the following sections.

In [Fig 2](#), we show comprehensive map of molecules that was manually assembled based on the published literature. Various entities of the network, e.g. genes, proteins and their modifications,



**Fig 1. Flowchart for generation of comprehensive map.**

doi:10.1371/journal.pone.0146759.g001



**Fig 2. A comprehensive map of obesity in human.** (Also see URL:<http://tinyurl.com/dykn8fd>).

doi:10.1371/journal.pone.0146759.g002

protein complexes are described using standard Systems Biology Markup Language (SBML) with the help of Cell Designer 4.1 software[35] and Systems Biology Graphical Notation (SBGN)[36] ([www.sbgn.org](http://www.sbgn.org)). The nodes (also known as species) represent molecules that participate in a given reaction. The edges represent reactions among nodes. The resulting network on obesity consists of 804 nodes (includes set A molecules as well as other genes/molecules interacting with set A) and 971 edges. These 804 nodes are categorized as 510 proteins, 115 genes, 1 ion, 3 drugs, 3 degraded molecules, 62 complexes, 23 RNA molecules, 83 simple molecules, 3 phenotypes and 1 unknown molecule (See Table F in [S2 File](#)). The reactions are categorised as follows: 848 state transitions, 33 transcriptions, 18 translations, 5 transports, 62 heterodimer association and 5 dissociations (Table F in [S2 File](#)). Supporting information for each interaction in the comprehensive map is documented separately in Table G in [S2 File](#).

### Linking lesser-studied/reported molecules with comprehensive map

There are several clinical conditions as well as syndromes, where obesity is one of the reported phenotype apart from other clinical features characteristic of that syndrome. These include Prader-Willi syndrome, Ulnar-mammary syndrome and Biemond 2 syndrome (See Table E in [S2 File](#)). Prader-Willi syndrome is characterised by hyperphagia, characteristic facial features, hypogonadism and short stature. This syndrome is caused by loss of genes imprinted on 15q11-q13 region such as gamma-aminobutyric acid (GABA) A receptor, gamma 3 (GABRG3), imprinted in Prader-Willi syndrome (IPW), small nucleolar RNA, C/D box 116 cluster (PWCR1), small nuclear ribonucleoprotein polypeptide N (SNRPN) and MAGE-like 2 (MAGEL2). We retrieved information for these genes from literature databases as well as from relevant pathways such as KEGG [37] and REACTOME [38]. Then, we aimed to find any



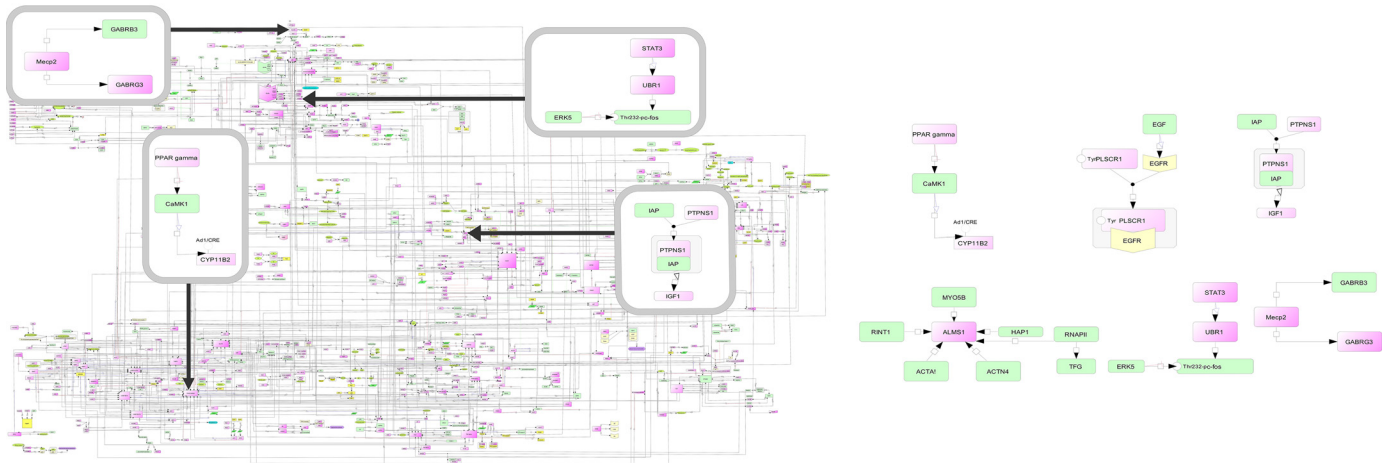
evidence of relationship between less-studied genes with obesity network molecules. After extensive manual screening, we were able to find one study [39] which links the **GABRG3** with methyl CpG binding protein 2 (MECP2) gene. The MECP2 gene is a part of module 1 of comprehensive network (See Fig 3). Encouraged by this result, we screened over 6000 abstracts representing 84 lesser-studied genes using our text-mining approach. These includes molecules such as, CYP11B2 (cytochrome P450, family 11, subfamily B, polypeptide 2), PLSCR1 (phospholipid scramblase 1), PTPNS1 (signal-regulatory protein alpha gene interactions), ALMS1 (alstrom syndrome 1), UBR1 (ubiquitin protein ligase E3 component n-recogin 1) and GABRG3 (gamma-aminobutyric acid A receptor, gamma 3). These efforts led to identification of several abstracts/studies through which we could link highly connected nodes with lesser-studied genes with the help of intermediary molecules. For example, we could identify that **CYP11B2** (molecule belonging to lesser studied group) expression and secretion is inhibited by peroxisome proliferator activator receptor gamma (PPAR  $\gamma$ ), a key molecule in adipocyte lineage [40] and a reported hub of our map. Additional details are given at our website <http://tinyurl.com/knnqsmm>. We also computed composite score of lesser studied (reported) genes and compared with hubs (well studied genes) (See File I in S1 File).

## Structure of the map

We used standard techniques to find structure in the constructed map [41]. The map has bow-tie architecture and resembles to alphabetical character "I". To facilitate map exploration, we divided our map into three regions: top, intervening (or central) and bottom. The prominent class of molecules present in the top region include neurotransmitters (catecholamine, dopamine, & serotonin), lipoproteins [lipoprotein lipase (LPL) & high density lipoprotein (HDL)], receptors [peroxisome proliferator-activated receptors delta  $\delta$  (PPAR  $\delta$ )], and cytokines [interleukin 6 (IL6)]. This indicates involvement of wide variety of molecules in obesity pathophysiology. The central part includes extensively reported molecules (leptin, ghrelin & insulin) along with less frequently reported molecules such as forkhead box A2 (FOXA2/HNF3B), pancreatic and duodenal homeobox 1 (PDX1) and lep-ob-Rb (leptin-leptin receptor) complex (See Table 1). The bottom region majorly consists of transcription factors and signaling molecules, inclusive of glucose transporter 4 (GLUT4), adiponectin (ADIPOQ), lipin1 (LPIN1), fatty acid binding protein 4 (FABP4), necdin, BMP delta-like 1 homolog (DLK1/PREF1), tumour necrosis factor alpha (TNF  $\alpha$ ) and PR domain containing 16 (PRDM16). In addition, several feedback loops connect top and bottom regions (highlighted in dark green colour in Fig 4).

## Representation of the map

Genes and proteins are represented by standard notations, whereas interactions are categorized as positive, negative, neutral and catalysis. A positive regulation is defined for set of molecules, in which the molecule's activity is stimulated by another molecule. In this context, authors frequently uses specific verbs such as stimulate, activate, induce, enhance, up-regulate and increase. Negative regulation is the inhibition of the neighbouring interacting molecule which is evident by verbs like inhibit, down-regulate, decrease, prevent, suppress and reduce. The edge representations include transcription, translation, association and dissociation using standard graphical notation. Apart from these, there are some reactions where a molecule regulates the reaction between other molecules, i.e. catalysis (See Table 2). The colour scheme and graphical representation is explained at website (<http://tinyurl.com/dykn8fd>) and File J in S1 File.



**Fig 3.** (A) Linking lesser studied (reported) genes (CYP11B2, UBR1, MECP2 and PTPNS1) with molecules of comprehensive map. (B) Examples of lesser studied gene interactions (CYP11B2, PLSCR1, PTPNS1, ALMS1, UBR1 and GABGR3).

doi:10.1371/journal.pone.0146759.g003

### Module generation—Reverse Engineering of the pathway

To understand a large network, a logical step is to divide the network into biologically meaningful smaller functional components [42]. This process is often termed as reverse engineering and several approaches have been described to identify modules. These range from spectral methods [43–44], methods that identify maximum flows or minimum cuts [45–46], heat kernels [47], betweenness centrality [48], seed node searches [49] e.g. MCODE in cytoscape [50], brute force methods [51] and weighted kernel k-means [52]. *Community structures* or *modules* are defined when a larger density of links exists within a specific part of the network than outside it [53]. We used different methods to identify community structures (modules) in obesity network. In addition, we clustered genes based upon tissue specific expression data. Since each method produces different results with some degree of overlap, we decided to integrate information to identify functionally meaningful modules (See File A in S1 File). Hence, the constructed network was divided into 5 modules based upon physiological processes and likely anatomical component (Table 3). In the following section, we attempt to relate modules with disease conditions (Fig 5).

**Module 1:** This module consists of highly connected nodes involved in neuro-hormonal signaling affecting energy homeostasis, hunger and mood. They include leptin, ghrelin and dopamine. Leptin is one of the highly studied molecules in obesity after insulin (present in 7.4% of total abstracts). Leptin acts as a satiety factor and its discovery has paved the way for the study of adipocyte derived factor in energy balance homeostasis. Further, the secretion of the leptin is directly proportional to amount of fat cells [54]. Recently, leptin replacement therapy has been proposed to treat obese individuals [55]. Frequent association of obesity with clinical depression can be explained by the impaired leptin activity in brain [56].

Ghrelin act as an endogenous ligand for growth hormone secretagogue receptor (GHSR). It has been reported to be involved in energy regulation and appetite signaling through activation of peptides, including AgRP, NPY and POMC [57]. Rise and fall in plasma ghrelin levels before and after food intake supports the hypothesis that ghrelin plays a physiological role in meal initiation in humans [58]. Ghrelin levels are altered in individuals suffering from Prader-Willi and Cushing’s syndrome [59]. A meta-analysis linked gastrointestinal hormones, ghrelin and obestatin levels with obesity [60].

**Table 1. Shows three components of the map: top, bottom and central component along with the molecules and their connectivity degree.**

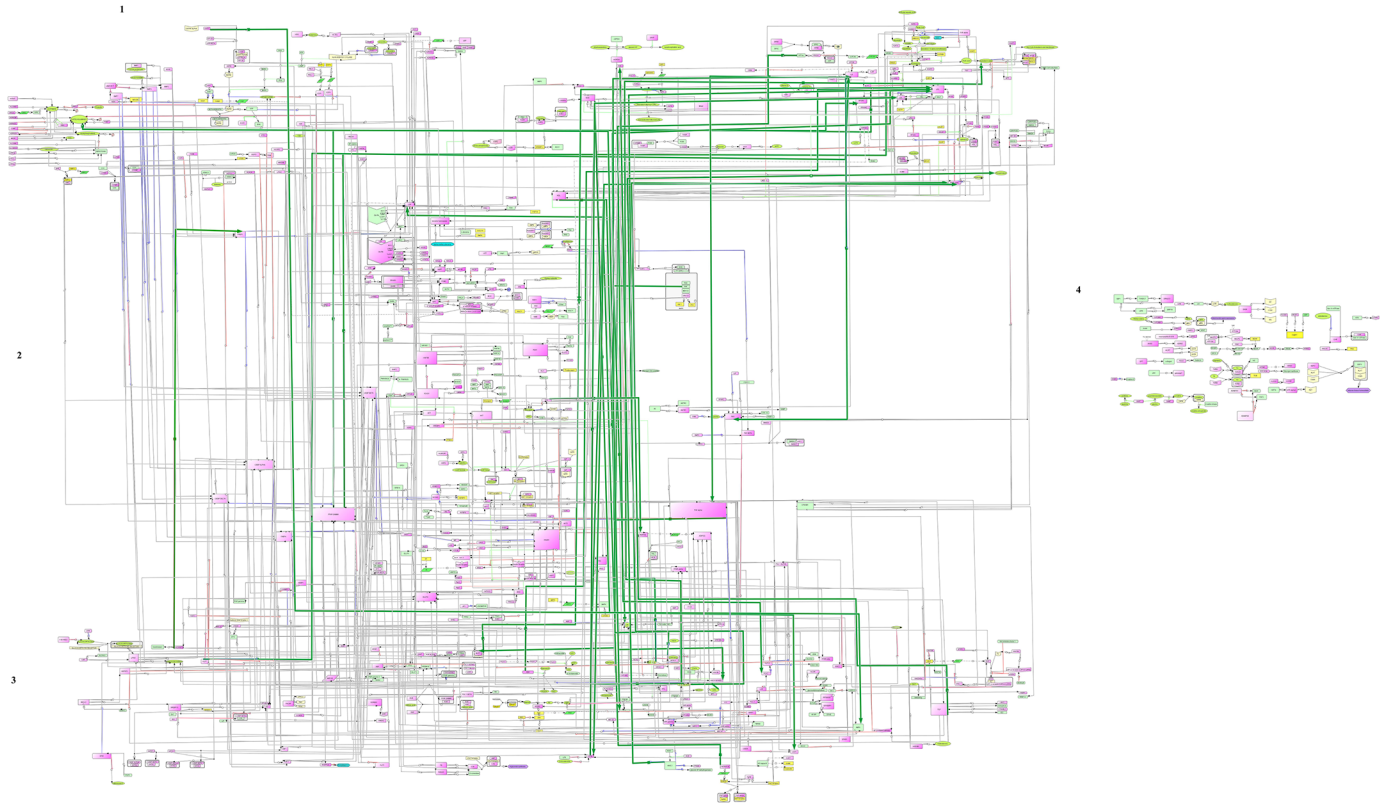
I SHAPE STRUCTURE			
CENTRAL COMPONENT		BOTTOM PART	
NODES	CONNECTIVITY DEGREE	NODES	CONNECTIVITY DEGREE
LEP	31	C/EBP ALPHA	24
OB-Rb-LEP	8	C/EBP BETA	18
OB-Rb	2	C/EBP DELTA	14
TYROSINE HYDROXYLASE	5	PREF1	11
GHRELIN-GHSR	8	NECDIN	7
AGRP	8	BMP7	9
POMC	7	BMP2	7
NPY	7	BMP4	7
ALPHA-MSH	6	LPIN1	8
PDX1	8	AGPAT2	5
HNF3B	7	BSCL2	4
FOXO1	7	SREBP1C	6
AKT	6	FABP4	12
SREBP1	6	GLUT4	18
PI3K	5	PPAR GAMMA	41
INSULIN	32	MIR103	6
IRS1	5	PPAR GAMMA 2	6
P38	11	PRDM16	7
<b>TOP PART</b>		ASP	5
DOPAMINE	9	TNF ALPHA	18
CATECHOLAMINES	8	PAI1	12
NOR-EPINEPHRINE	6	PGC 1 ALPHA	12
TRYPTOPHAN	3	ERR GAMMA	7
SEROTONIN	6	PPAR ALPHA	9
CCK-CCKAR	3	ADIPOQ	8
EPINEPHRINE	3	NRF2	3
DRAJC3	3	ERR ALPHA	7
ALPHA ADENYL CYCLASE	4	NRIP1	5
SCD	9	TCF	12
MLXIPL	4	CTNNB1	12
LPL	13	TCF4	4
HDL	4	AMPK	8
FXR ALPHA	5		
CHOLESTROL	7		
IL6	11		
CHOLESTROL ESTER	7		
PPAR DELTA	6		

doi:10.1371/journal.pone.0146759.t001

The importance of dopamine signaling in obesity has been demonstrated by the alteration of dopamine receptor levels with changes in body mass index (BMI) [61]. Apart from these, several other molecules have been reported in context of obesity; therefore, we have described their roles at our website: <http://tinyurl.com/kazahj6>.

**Module 2:** Obesity is a major risk factor for non-insulin dependent diabetes mellitus (NIDDM) [62]. Insulin is a central molecule in pathophysiology of type 2 diabetes and also





**Fig 4. I-shape structure of the network consisting of central component connected with top and bottom regions.** The feed-back loops are highlighted in dark green colour.

doi:10.1371/journal.pone.0146759.g004

appears in large number of abstracts related to obesity (23,165 abstracts; 24% of total dataset) in humans. Module 2 primarily encapsulate insulin and its interactions with other molecules, for instance, apolipoprotein A-V (APOA5), forkhead box C2 (FOXC2), macrophage migration inhibitory factor (MIF), uncoupling protein (UCP) and v-akt murine thymoma viral oncogene homolog 2 (AKT2). This module builds a link between tightly coupled clinical conditions- obesity and type 2 diabetes.

**Module 3:** Lipid storage and metabolism is affected frequently in obese patients leading to dyslipidemia, exposing them to cardiovascular risks [63] and atherosclerosis [64]. The third module maps interactions, catalysis and processing of molecules involved in lipid metabolism, including acetyl CoA, aspartate, mevalonate, cholesterol, cholic acid, and diacylglycerol.

**Module 4:** It is the largest module in the network and majorly consists of transcription factors involved in adipose tissue differentiation and other biological activities in humans. The interactions are dominated by molecules like peroxisome proliferator-activated receptors-PPAR ( $\alpha$ ,  $\beta$ ,  $\gamma$ ) and CCAAT/enhancer-binding proteins-C/EBP ( $\alpha$ ,  $\beta$ ,  $\gamma$ ). The molecules such as PPAR  $\gamma$  (with 41 edges) provide indirect connections with lesser studied genes/molecules reported in context of obesity. This module is divided into another sub-module labelled as “4A” to incorporate set of molecules distinct from transcription factors.

**Module 4A:** Though the Wnt pathway has been shown to play a major role in embryogenesis and some of the cancers, it has also emerged as an important regulator of adipocyte differentiation [65]. In addition, recent evidence of obesity treatment using traditional herbal medicine,

**Table 2. Shows types of interaction, example of verbs, representative sentences and references.**

Types of interaction with verbs	Sentences	References
<b>Positive</b>		
Activate	Leptin in central nervous system exhibit their action by binding to ob-Rb, a long isoform of leptin receptor. The binding leads to the <b>activation</b> of four tyrosine residues by phosphorylation and a Box 1 region in ob-Rb.	Frühbeck,2006
Positive	The STAT3 exhibits a <b>positive</b> regulation on POMC [pro-opio melanocortin], which are the major molecules in energy regulation	Frühbeck, 2006
Induce	C/EBP $\beta$ and C/EBP $\delta$ <b>induces</b> kruppel-like transcription factor [KLF] 5.	A. Schäffler et al, 2006
Stimulate	E2F1 <b>stimulate</b> PPAR $\gamma$ during early stages and E2F4 inhibit it in later stages.	Fajas, L. et al, 2002
Up-regulate	In adipocytes, SORBS1 is <b>upregulated</b> by thiazolidinedione [TZD] through PPAR $\gamma$ and also interact with insulin receptor.	Wen-Hsing Lin et al, 2001
Increase	PRDM16, which leads to <b>increase</b> in expression of UCP1, PGC $\alpha$ and PGC $\beta$ during mature adipocyte development	Yu-Hua Tseng et al, 2008
Enhance	Urocortin <b>enhances</b> Leptin-Induced Stat1 Activation, Mediated by Either CRHR1 or CRHR2.	Weihong Pan et al,2007
Require	We report that exercise and recombinant IL-6 requires IL-10 expression to suppress hyperphagia-related obesity.	Eduardo R. Ropelle et al, 2010
Bind	Follistatin binds to myostatin but also binds to and inhibits other members of the TGF-beta superfamily, notably activins.	Nakatani M et al, 2008
Trigger	AMPK activated through leptin and external stimuli, <b>triggers</b> myocytes enhancing factor 2 [MEF2] [A & D] which leads to the expression of GLUT4.	Farid F. Chehab, 2008
<b>Negative</b>		
Inhibit	PPAR $\gamma$ <b>inhibit</b> the expression of aromatase by CYP19A1 gene.	Reposo Ramirez-Lorca et al, 2007
Negative	LXR [liver X receptor], a transcription factor take part in adipogenesis by <b>negatively</b> regulating PGC-1, glucose-6 phosphatase, phosphoenol pyruvate carboxykinase [PEPCK/PCK2].	Bryan A. Laffitte et al, 2003
Prevent	PON3 in plasma individually <b>prevents</b> atherosclerosis by inhibiting phospholipid stimulation of monocyte chemotactic protein [MCP]	Diana M. Shih et al, 2007
Decrease	PPAR $\delta$ <b>decreases</b> LDL and triglycerides.	Markku Vañntinen et al, 2005
Reduce	In the presence of ID3, the binding of E47 and SREBP1 is <b>reduced</b> , thereby transcriptional activity of ADIPOQ is controlled.	Curt D. Sigmund et al, 2008
Suppress	In conclusion, disruption of neuromedin B receptor did not interfere with the sensitivity of thyroid hormone-mediated <b>suppression</b> of TSH release, but impaired the ability of thyrotroph to increase serum TSH in hypothyroidism, which highlights the importance of NB in modulating the set point of the hypothalamus–pituitary–thyroid axis at hypothyroidism.	Karen J Oliveira et al,2008
Down-regulate	We demonstrate that activation of LXR in the liver leads to the induction of glucokinase expression and to the <b>down-regulation</b> of peroxisome proliferator-activated receptor $\gamma$ coactivator-1 $\alpha$ [PGC-1] and genes involved in gluconeogenesis.	Bryan A. Laffitte et al, 2003
<b>Neutral</b>		
Regulate	The cyclic AMP responsive element–binding protein-1 [Creb1]- <b>regulated</b> transcription coactivator-1 [Crtc1] is required for energy regulation	Judith Y Altarejos et al, 2008.
Correlate	In girls, visfatin <b>correlated</b> with leptin, $r = 0.40$ , $P = 0.009$ , and thiols, $r = -0.36$ , $P = 0.009$ , which explained 24% in visfatin variability.	Krzystek-korpacka M et al, 2011
Elicit	At basal glucose, GIP does not <b>elicit</b> insulin release.	Pfeiffer AF et al, 2010
Influence	An exciting new report describes that leptin can <b>influence</b> insulin release by osteoclastin, a hormone produced by osteoblasts.	Rashmi C and Rodger AL, 2011
Associate	Leptin was <b>associated</b> with insulin, insulin resistance, and body composition parameters, body mass index, basal metabolic rate, body weight, %fat, and fat mass, in participants, with or without T2DM in both genders.	Gulturk S et al, 2008
Affect	In conclusion, HSL <b>affects</b> insulin secretary capacity especially in the setting of obesity.	Sekiya M et al, 2009
Contribute	FOXO1 may <b>contribute</b> to enhanced ADIPOR1, but not ADIPOR2 transcription in IR.	Felder TK et al, 2010

(Continued)

Table 2. (Continued)

Types of interaction with verbs	Sentences	References
<b>Catalysis</b>		
Catalysis	Tyrosine hydroxylase [TH] <b>catalysis</b> the conversion of L-tyrosine to DOPA which in-turn is regulated by SPOCK1, fgf14 [fibroblast growth factor-14], Cxcl14, BMP6 [bone morphogenetic protein 6]	C. Vadasza, 2007

doi:10.1371/journal.pone.0146759.t002

SH21B, has indicated about anti-adipogenic mechanism mediated by Wnt-β catenin signaling [66].

**Module 5:** The last module contains information about disjoint set of genes/proteins involved in obesity which are difficult to categorize due to inadequate information.

### Quantitative Analysis

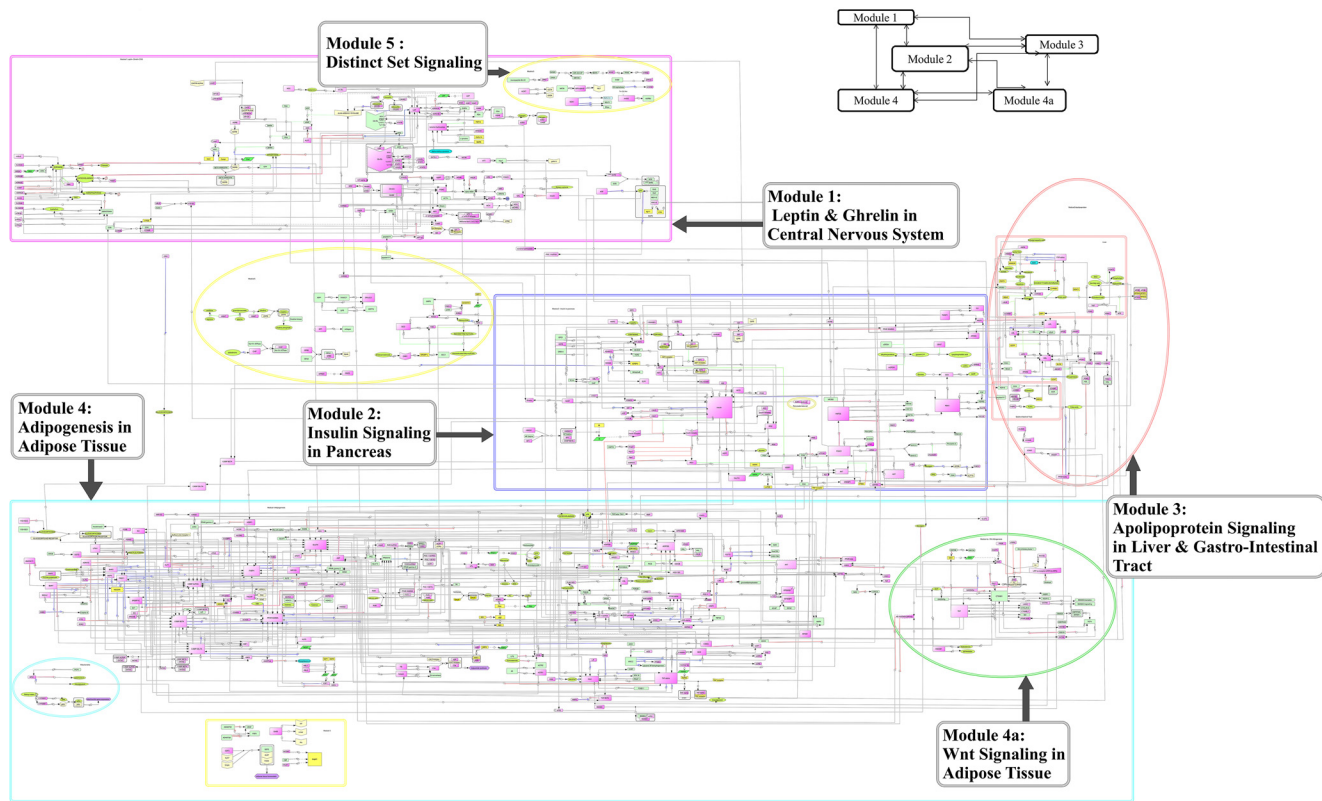
To understand the properties of constructed network, we computed several topological parameters as described below (See File B in [S1 File](#) for detailed information).

- Degree distribution parameter:** We found that the several number of connections follow power laws that indicates scale-free pattern of connectivity ( $\gamma_{in}$ : in-degree parameter as 2.19 and  $\gamma_{out}$ : out-degree parameter as 2.11). The scale free behaviour is also observed in constituent modules suggesting preferential attachments and hubs in the network (See [Table 4](#)).

Table 3. Describes information of five modules obtained from the network. The columns show pre-dominant hub, likely anatomical component and physiological process with the connectivity degree of major molecules.

Modules	Pre-dominant Molecule	Likely Anatomical Component	Physiological Process	Major Molecules with their degree of connectivity
Module 1	Leptin	Central Nervous System	Satiety, Appetite, Energy expenditure	Leptin [LEP]- 14, Dopamine- 9, Catecholamine- 8, Ghrelin- Growth Hormone Segretagogue Receptor [GH-GHSR]-5, Agouti-related Protein [AGRP]- 8, Pro-opiomelanocortin [POMC]- 6, NeuroPeptide Y [NPY]- 7, Tyrosine Hydroxylase-4, Adrenergic Receptor Alpha 1B [ADRA1B]-3
Module 2	Insulin	Pancreas	Glucose metabolism	Insulin [INS]- 21, Forkhead box A2 [HNF3B]- 6, Pancreatic and Duodenal Homeobox 1 [PDX1]- 7, Forkhead box O1 [FOXO1]- 6, Interleukin 6 [IL6]- 5 Leptin [LEP]- 5, Insulin receptor substrate 1 [IRS1]- 6, Insulin receptor substrate 2 [IRS2]- 5, Insulin receptor- 5
Module 3	High Density lipoprotein	Liver & Gastro-Intestinal Tract	Fatty acid regulation and metabolism, Base-line glucose control.	Lipoprotein Lipase [LPL]- 6, High Density Lipoprotein- 12, Low Density Lipoprotein- 6, FXR alpha- 5, Cholesterol- 6
Module 4	PPAR gamma	Adipose Tissue	Store fat derived from the diet and liver metabolism or degrades stored fat to supply fatty acids and glycerol to the circulation.	Peroxisome proliferator-activated receptor PPAR alpha- 9, PPAR gamma- 41, PPAR beta- 3, PPAR gamma 2-6, PGC1alpha -12 CCAAT/enhancer binding protein C/EBP alpha- 24, C/EBP beta- 18 C/EBP delta- 14, Glucose Transporter 4-9, Tumor Necrosis Factor Alpha-18, Adiponectin [ADIPOQ]- 8, PR domain containing 16-7, PREF1-11, Necdin- 7 Lipin1- 8, Sterol regulatory element binding transcription factor 1 SREBP1c- 6
Module 4a	CTNNB1	Adipose Tissue	Inhibitor of Adipogenesis	Catenin [cadherin-associated protein], beta 1 CTNNB1-12 HNF1 homeobox A [TCF]- 9
Module 5	Distinct Sets Signalling	-----	-----	-----

doi:10.1371/journal.pone.0146759.t003



**Fig 5. Modules of the comprehensive map.**

doi:10.1371/journal.pone.0146759.g005

2. **Clustering Coefficient:** Our network and its constituent modules show clustering coefficient values and average clustering coefficient [67] close to 0 suggesting tree-like structure [68].
3. **Average shortest path length** value was found to be 15.85 for comprehensive network supporting scale free nature of the graph [69].

### Randomization of Constructed Network

We constructed null models (control) and compared the properties of comprehensive network with null models [68, 70]. The protocol is described as following: In a true network gene A (leptin) binds with gene B (leptin receptor) to perform a function X (i.e. leptin act as satiety factor and exhibits its action by binding to leptin receptor) in cell.

Null model 1- In this model, we randomised the edges but kept the node labels and their degrees intact. For example, the connection (edge) between gene A (leptin) and gene B (leptin receptor) is deleted. A new connection is established between gene A and gene C (any other gene of the network except leptin receptor) so as to disrupt the function X.

Null Model 2- We shuffle the positions of nodes by keeping the global degree distribution of the comprehensive map intact.

Null Model 3- This is generated by shuffling both the position of nodes as well as their edges (See File C in [S1 File](#)).

Null Model 4- We construct the network with same number of edges and nodes using methods proposed by Erdos-Renyi [71], Watts-Strogatz [72] and Barabasi-Albert [73]. To see the

**Table 4. Topological analysis of the comprehensive map using Network analyzer and Gephi.**

S.No.	Parameters	Network Analyzer	Gephi
1	Nodes	1799	1799
2	Degree Distribution		
	In-degree		
	Y	2.193	-
	R <sup>2</sup>	0.866	-
	Out-degree		
	Y	2.116	-
	R <sup>2</sup>	0.914	-
3	Clustering Co-efficient	0	-
4	Connected Components		
	Weakly Connected	35	35
	Strongly Connected	-	1398
5	Diameter	46	46
6	Average Shortest Path-length	15.86	15.86
7	Average number of neighbours	2.37	
8	Network Density	0.0	0.001
9	Average Degree	-	1.192
10	Average Weighed Degree	-	1.6
11	Modularity	-	0.875
12	No. of Clusters	-	62

doi:10.1371/journal.pone.0146759.t004

effect of properties on size of the network, we construct networks with node numbers from 100 to 1000. Firstly, we use method proposed by Erdos-Renyi to construct a random graph of N nodes connected with n edges, which are chosen randomly from N(N-1)/2 possible edges and are not scale-free [71]. Secondly, a control network was generated through Watts-Strogatz model (1998) [72], where in random graph is produced with small-world properties, including short average path length and high clustering. Thirdly, in Barabasi-Albert model [73], the generation of random graph is based on the connected seed network of s nodes. Remaining nodes (n-s) are added one at a time, and connected to existing nodes (m) randomly. The resulting network is found to follow power-law degree distribution.

In addition, we generated randomized networks using random network module of cytoscape. The obesity network (true network) exhibit different properties when compared to 18 control randomized networks obtained by shuffling the obesity network associations while keeping the degree distribution of nodes fixed (Fig G and Fig H in S1 File). We find that clustering coefficient increases from 0 (in true network) to 0.00201 (randomized network with 30000 shuffling. See Fig G in S1 File). This pattern is reversed in case of mean shortest path, which reduced from 18 to 11 units (See Fig H in S1 File). We have also enclosed additional information for results generated during shuffling procedure in the Table H in S2 File and website in S3 File.

### Robustness of Network

To see the robustness of network and its dependence on failure of a particular node, we randomly deleted nodes and computed properties for the remaining network. There are several indexes of network centrality such as degree, eccentricity, closeness, betweenness, stress, centroid and radiality which allow quantifying the topological relevance of single nodes in a



network. Recently introduced parameters such as node interference and robustness were also included in the analysis. These parameters measure the relative importance of given node in context of network [74]. It was also shown in the past that the hubs (nodes with high degrees) play important roles in maintaining structural integrity of networks against failures and attacks, [75] in spreading phenomenon [76] and in synchronisation [77]. Since, obesity network shows scale free structure with presence of hubs, we started our deletion experiments by sequential deletion of hub nodes to see the effect on network robustness. This was achieved by removing a node and calculating the interference on the centrality of the remaining nodes using centiscape plugin of cytoscape. We find that removal of hubs alone or in combination impact the network tremendously. We find that various critical properties of network changes to significant extent. For example, betweenness of nodes of original network (Mean = 10738.4; Var. = 3.3E8) were significantly different when compared with networks obtained after deleting all hubs (Mean = 13675.4; Var. = 1.1E9) computed through paired t test ( $P < 0.05$ ). When we randomly deleted any node (not hub), the changes were not significant in the parameters (See Table I in [S2 File](#) and Folder: Deletion experiment on website (A) in [S3 File](#)).

### Gene Ontology Analysis (GOA)

To understand the biological processes present in large dataset of obesity genes, we used BiNGO [78] and Network Ontology Analysis (NOA) [79]. We observed that adipocyte specific functions, including response to nutrient level were represented by 48 molecules (10% of total dataset, p value:  $2.49 \times 10^{-25}$ ), regulation of lipid metabolic process was seen with respect to 47 molecules (9.9% of total dataset, p value:  $2.44 \times 10^{-37}$ ), carbohydrate metabolic process (38 molecules; 8%, p value:  $9.5 \times 10^{-07}$ ), lipid localization (36 molecules; 7.6%, p value:  $3.75 \times 10^{-22}$ ), lipid biosynthetic process (35 molecules; 7.3%, p value:  $5.42 \times 10^{-10}$ ), feeding behaviour (31 molecules; 6.5%, p value:  $8.0 \times 10^{-30}$ ), and response to nutrient by 31 molecules (6.5%, p value:  $2.49 \times 10^{-25}$ ). In biological process, the sub-category- cellular process comprises 80.9% of the genes of our dataset, which include cell communication (83 molecules; 22.6% of total dataset, p value:  $1.06 \times 10^{-22}$ ), regulation of gene expression (152 molecules; 41.5%, p value:  $4.43 \times 10^{-12}$ ) and regulation of programmed cell death (72 molecules; 19.6%, p value:  $6.89 \times 10^{-15}$ ). Similarly in molecular function, 92.9% genes are involved in binding activity and in cellular process/location- 69.5% of the genes are found to be present in intracellular section of cell which primarily includes nucleus (165 molecules; 59.1% of total dataset, p value:  $1.03 \times 10^{-02}$ ) and endoplasmic reticulum (50 molecules; 17.9%, p value:  $4.33 \times 10^{-05}$ ) (See Table J in [S2 File](#)).

### Mapping of Microarray Data

The microarray data was obtained from Gene Expression Atlas [80] using search term “obesity and homo sapiens” from the URL (<http://www-test.ebi.ac.uk/gxa/>). The gene list was obtained for three possible conditions: up-regulation, down regulation and non-differentially expressed. We selected 3,485 genes reported to be up-regulated in obesity and labelled them as set ‘U’. Subsequently, we found 2,135 genes (labelled as D) as down-regulated group and a very large number of genes (2,91,407) as non-differentially expressed (NDE). After removal of redundancy, we obtained 1,340 molecules as up-regulated (U), 918 molecules as down-regulated (D), and 38,434 molecules as non-differentially expressed (NDE) molecules as a filtered sets.

Thereafter, we compared filtered dataset obtained from microarray database with our list. Based upon comparisons, we found that 27 genes (obtained from deep curation approach (DC)) are up-regulated in obesity whereas 24 genes show down-regulation and large numbers of genes did not show any change in expression or information is not available in the database. Using gene ontology analysis, it was revealed that most of the up-regulated genes are involved



in protein binding and down-regulated group are involved in steroid binding activity (See File D in [S1 File](#)).

Since, we could not map large number of genes; we attempted to find expression data of obesity genes in GEO (<http://www.ncbi.nlm.nih.gov/geo/profiles>) microarray database. We found that 34.5% of genes (obtained from text mining (TM) approach) are up regulated whereas 27.58% are down regulated.

## Applications of Obesity Network-Implications in therapeutics

We used orlistat (tetrahydrolipstatine, an FDA approved drug for treatment of obesity) to dock against the molecules listed in our network using our in-house docking pipeline “Docoviz”. We observe that orlistat not only binds to fatty acid synthase (FASN) ( $\Delta E = -13.7$  Kcal/mol; experimentally known target) but also binds to several other molecules in the obesity network. To check whether orlistat produces its clinical effect (of weight reduction) possibly due to preferential binding to several molecules listed in the obesity network (N) than any other part of proteome, we created a dataset of 24,000 known human protein structures (P) and docked orlistat against them. In addition, we created datasets of randomly selected protein structures from P labelled as P1, P2. . . Pn as controls. We also used Alzheimer disease network molecules [81] as an additional control (D). We observed that the distribution of binding energies obtained from controls (P1, P2, P3. . . Pn) and Alzheimer disease network(D) is significantly different from test dataset(N) (P value  $<0.05$ , Welch T test).

In another experiment, we docked drugs (which do not have effect on obesity) against the obesity network proteins. For instance, we used Acetylsalicylic acid (selected randomly; anti-inflammatory medicine) to dock against the obesity network proteins. Apart from that, we used drugs, showing comparable tanimoto co-efficient to orlistat, such as 3-Carboxy-N,N, N-Trimethyl-2-(Octanoyloxy) Propan-1-Aminium (Tc Value: 0.68) and 6-DeoxyerythronolideB (Tc Value: 0.6) to ascertain binding energy profiles in the obesity network. We detected that the binding energy profiles of the above mentioned drugs against the obesity network proteins are different from that of orlistat (P value  $<0.05$ , Student’s T test).

Orlistat is known to produce several side-effects namely acne, respiratory tract infection, urinary tract infection and nausea, possibly due to binding to off targets perturbing unrelated pathway. Using text mining systems and manual screening, we obtained list of molecules implicated in the side effects/diseases related to orlistat. On comparison, we found that several molecules are common in obesity network and acne (14 molecules; 2.7% of total dataset), providing a possible clue for causation of acne in patients taking orlistat during obesity treatment. Likewise, sibutramine (antidepressant and anorexigenic drug) was withdrawn due to adverse effects such as agitation, fever, vomiting, diarrhoea, loss of coordination, and dilated pupils. Using our map, we could link the side effects of sibutramine with disease networks. To illustrate, symptoms such as nausea, vomiting and depression are likely to be produced due to binding of sibutramine to targets such as SLC6A3 and SLC6A4 and subsequent perturbation of pathway involving HTR2C (anxiety), HTR2A (anxiety), DRD2 (nausea and vomiting), COMT (nausea and vomiting), and MAOA (depression) (File E in [S1 File](#)).

## Discussion

This work shows a new approach of combining data from heterogeneous databases including literature, structure and microarrays to construct disease networks and attempt to explain therapeutics of a drug molecule in context of networks. Our methods are generic, web enabled and open in nature to build rich networks. Each entity i.e. node or edge has been hyper-linked to its

source (research papers) so as to maintain transparency in the system for users to evaluate and improve the system in a collaborative fashion.

Network targeting involves activity of a compound across multiple pathways which might be necessary to effectively stop neoplasm and pathogens, but can also produce side effects by targeting undesirable proteins [82]. Very few large scale docking studies have been conducted in the past (Gao *et al.* used ~1,100 targets [83]; Hui-fang *et al.*, used 1,714 targets [84]; [85]). Here, we performed docking of orlistat with obesity network proteins as well as with whole human proteome (>24000 proteins) as a test example. Based upon our predictions, we propose that a given drug (orlistat) not only bind to its known target (FASN;  $\Delta E = -13.6$  Kcal/mol) but also to several other targets in the network with varying degree of binding energies. This propensity of binding of drug within the target network (obesity) is different from binding with any other disease network or network randomly drawn from human proteome. Further, we also observe that the therapeutically unrelated drugs for a given clinical condition (“Acetylsalicylic acid in obesity”) show different binding patterns to network proteins. These results contribute to emerging concepts of network pharmacology [82] and chemigenomics [86] to develop safer, cheaper and effective medicines. The possible limitation of this approach is non-specific or random binding of ligand to many of the protein targets.

Real world networks including biological networks are characterised by presence of few highly connected nodes known as hubs and they tend to show non-Poisson degree distribution. Evidence shows that hub proteins are encoded by essential genes [87] that seem to be older, evolve slowly and their deletion affect a large number of nodes as compared to non-hub nodes [88–90]. Therefore, different studies have attempted to associate hub proteins to disease genes. Some studies support this hypothesis, whereas few studies contradict this hypothesis [90–92]. Our network shows hub based architecture with select set of nodes occupying most of the connections- leptin, insulin and PPAR gamma. Most of these genes likely to be essential in nature, whereas some of the recently reported candidate genes are present in periphery in our map, e.g. fat mass and obesity associated (FTO) gene. It may be inferred that the obesity pathophysiology is primarily influenced by interactions of essential genes, therefore obesity could be considered as a system level adaptation toward chronic nutritional over intake and other causative factors.

We compared our network with previously published dataset, including Kitano *et al.*, 2004 [41], Logsdon *et al.*, 2012 [93] and found several of our network molecules present in these datasets (See File F in [S1 File](#)). Various population wide studies have indicated that hypertension is a predominant clinical condition affecting over 40% of obese people (BMI > 30) [94], whereas type II diabetes mellitus affects 40–60% of obese people [95]. Using text mining approaches, we found that there is a significant overlap between molecules implicated in obesity and its associated disorders such as diabetes or hypertension. This overlap is less when molecules implicated in obesity are compared to molecules implicated in unrelated disease group e.g. asthma, urticaria and ataxia.

Considering wide variety of factors affecting the obesity pathophysiology, we believe that obesity comprehensive map will act as a platform to integrate information derived from gene expression experiments, protein-protein interaction data, drug information, clinical data, metagenomic and pharmacogenomic information. It will be interesting to understand how this network evolves temporally in a lifespan of a given individual(s) from lean state to obese state. What modules or links get formed or abolished during the process? It can also act as a system where new drugs may be tested against disease networks to predict their therapeutics or side-effects.

## Material and Methods

### (A) Retrieval of Literature Data

We screened each research article manually and highlighted text for the name of molecules as well as their interactions. We also used information provided in human obesity gene map database 2005 update [30] and GenMapp (<http://www.genmapp.org/default.html>). The abstracts having the term “obesity” and “human” were downloaded from PubMed using RefNavigator (version 2.0). We obtained 96,219 abstracts on obesity in human till December 2012 (See Folder 2 available at website (A) in [S3 File](#)). We used perl scripts to parse additional information which includes authors’ names, affiliations, journal name & year of publication. Each abstract was processed and unique id was assigned using perl scripts.

### (B) Determining True Positives and False Positives

Researchers have used several approaches to link genes with complex traits such as obesity. Primarily, linkage analysis and association studies have been used to find the variants that affect obesity. In addition, animal models also provide list of candidate genes through linkage studies, expression profiling, and transgenic strains. The techniques such as expression analysis and protein interaction studies also identify candidate genes for obesity. Given the wide variety of available experimental techniques, we grouped these studies (evidences) into various categories and provided a numerical code to each of them (See Table B in [S2 File](#)). Next, we label each gene with a numeric code for better data management.

A gene is defined as true positive example, when we have enough evidence to link a gene with a disease. For example, Leptin (Lep) deficiency is linked with intractable form of obesity (Uniprot Id—P41159; OMIM ID- 614962). As a rule of thumb, we labelled genes with high confidence when many independent research studies published in high impact journals with sufficient citations support that link. Since, each gene has different types of experimental evidences ranging from mutation studies, animal studies, genome wide association linkage studies and clinical studies. We grouped these evidences into various categories and provided a numerical code (See Table B in [S2 File](#)). The false positives are those gene examples which matched common English words used in sentences, abbreviations of organizations, and author names. They also include examples which occurred in abstracts but rejected during manual screening due to lack of clear evidence.

### (C) Hybrid approach

Deep-curation approach (DC) is defined as screening of literature data by experts whereas text-mining systems (TM) sift through publication data for the occurrence of the genes and their interactions using computational software and predictive algorithms at large scale. Though, text mining systems are fast, but they suffer from several problems limiting their use. For example, consider a representative statement from a research article [54], “the binding of the SH2 domain of SH2B1 to phospho-Tyr, 813, in JAK2 enhanced leptin induction of JAK2 activity”. Here, different text mining tools will report—“Jak2 enhanced leptin”. This is considered to be a positive interaction but the real meaning is leptin increases JAK2 activity upon binding of SH2 domain to JAK2. Due to these constraints, text mining systems are not considered robust enough to resolve numerous problems warranting the need for deep-curation approach. Our TM approach is formulated as following-:

- (i). Let  $W$  be the set of all the genes and their synonyms in human that may occur at least once in the set of abstracts labelled as  $A$ . The  $W$  is represented as a matrix where each row represents a gene ( $w_i$ ) and its synonyms. The synonyms and approved symbols for

each gene are shown in tab separated format in a text file where notation “ $w_{ij}$ ” is designated for them.

- (ii). A separate matrix ( $M$ ) is constituted for storing frequencies of genes, listed in  $W$ . It contains genes ( $w_i$ ) in the first column and their respective counts ( $c_k$ ) in second column. For example,  $w_2$  represents the gene LEP, having a gene count ( $c_2$ ) of 7,159 in the PubMed abstracts (1960–2012 December).
- (iii). We also define  $N$  as the gene co-occurrence matrix. Each entity of this matrix is described as  $N_{xyz}$  to store information extracted from research articles. This is composed of three units:  $N_x$ ,  $N_y$  and  $N_z$ .  $N_x$  capture first instance of gene encountered in the sentence whereas  $N_z$  keeps the next instance of gene and  $N_y$  stores intermediate set of words. To illustrate, consider a statement, “**Insulin** is known to increase expression of the *ob* gene product **leptin** in adipose tissue”. Here, insulin and leptin are labeled as gene pair having 10 intermediate words between them. Therefore, “insulin” will be  $N_x$ ; “leptin” will be  $N_z$  and “is known to increase expression of the *ob* gene product” is  $N_y$ .

$$MatrixN = \begin{matrix} & \begin{matrix} Gene1 & InterveningWords & Gene3 \end{matrix} \\ \begin{matrix} N1 \\ N3 \\ N4 \end{matrix} & \begin{matrix} N2 \\ N5 \\ N6 \end{matrix} & \begin{matrix} N3 \\ N1 \\ N1 \end{matrix} \end{matrix}$$

We extract gene pairs from the abstracts and full length articles and compute their frequencies. We also build frequency distribution of intermediate words ( $N_y$ ) useful for building dictionary for subsequent natural language processing. This dataset is also useful for training of machine learning systems such as hidden markov models and support vector machines (manuscript in preparation) as well as manual curation.

- (iv). Parser is a set of dictionaries that are built for various types of interactions, tenses and negations. We curated data of 300 research articles to identify the most frequently used words to represent interactions namely, positive, negative and neutral. We use these dictionaries to label interactions by building a matrix  $O$ . In matrix  $O$ ,  $O_{xyz}$  represent the data structure where the gene  $O_x$  (insulin) is followed by gene  $O_z$  (leptin) with their type of interaction,  $O_y$  (positive). This is processed for graphical-view using GraphViz (Version 2.28). The detailed example (tutorial) of TM approach is provided in a S7 in [S1 File](#).

### Text-Mining Approach Algorithm

```

Let abstracts = A;
Let genes = W; // 35,000 Genes in human & its synonym
Let gene count matrix = M;
Let co-occurrence matrix = N;
Let NLP matrix = O;
for i = 1 to n do // 'i' is a row representing a gene in W
for j = 1 to n do // 'j' is a column representing a gene name, symbol in W
    Let ck = 0; // Initializing the count of a gene 'i' in abstracts A as 0
if i, j ∈ A then
    write i to M; // write the gene 'i' in gene count matrix M
    ck = ck + 1;
append ck to M; // The gene 'i' is appended with its count ck in M

```

```

        next; // Search for the next gene
    read gene  $x \in M$ ; //Reading the gene  $x$  from gene count matrix  $M$ 
    read gene  $z \in W$ ; //Reading the gene  $z$  from the dictionary  $W$ 
    gene-pair  $N_{x,z}$ ;
for  $x = 1$  to  $n$  do //  $x$  represents the first gene of a gene-pair in  $M$ 
    for  $z = 1$  to  $n$  do //  $z$  represents the second gene of a gene-pair in  $W$ 
        Let  $N_{x,z} = 0$ ; // Initializing the count of gene-pair  $x, z$  as 0
        Let  $y = 0$ ; //  $y$  is words between gene-pair initialized as 0
        for gene  $x \in A$  do
            for gene  $z \in A$  do
                if  $x$  then  $z$  then
                    read  $y$  in  $A$ 
                    if length  $y > 3$ ;
                        write  $x, y, z$  to  $N$ ; //  $N_{x,y,z}$  is a co-occurrence matrix
                         $N_{x,z} = N_{x,z} + 1$ ; // Total occurrence of a gene pair
                        append  $N_{x,z}$  to  $N$ 
                next; // Search for the next gene-pair
Let NLP Parser =  $P$ ; //Set of Dictionaries  $P$ 
    Let Interaction verb dictionary =  $P_a$ ; //Sub-dictionary in  $P$ 
    Let tenses =  $P_b$ ; //Sub-dictionary in  $P$ 
    Let negations =  $P_c$ ; //Sub-dictionary in  $P$ 
    for  $P_a \in N$  // Search for interaction verb in  $N$ 
    for  $P_b \in N$  // Search for tenses in  $N$ 
    for  $P_c \in N$  // Search for negations in  $N$ 
        write  $O$ ; //  $O$  is a NLP matrix

```

#### (D) Comprehensive Map Construction

The comprehensive map of molecules in obesity was constructed using Cell Designer software [35]. Cell Designer support systems biology graphical notation (SBGN) and provides various functions to the users to represent molecular entities, including gene, protein, and RNA as well as edge notations-transcription, translation, inhibition and stimulation. The activity as well as modulation in the molecule can also be represented. The constructed map can be exported as systems biology mark-up language (SBML) format, preferred for computational models of biological processes.

#### (E) Module Generation

Reverse engineering of the comprehensive map was conducted using tools and methods mentioned in A File in [S1 File](#).

#### (E) Random Model Generation

Random models of the comprehensive map were generated by two approaches: Firstly, by Degree Preserving Random Shuffle using Network Analyzer Tool [96] and secondly, by applying Scale-free random graph (a cytoscape plug-in Random Networks). We also used perl scripts developed in-house for randomisation process.

#### (F) Comprehensive Map Analysis

1. The topological analysis was performed through graph based algorithms such as NetworkAnalyzer [97] and Gephi (<https://gephi.org/>).

- The gene ontology (GO) analysis was carried out for three categories: molecular function, biological process and cellular component using BiNGO [78] and Network Ontology Analysis (NOA) [79].
- The identification of protein targets of drugs, particularly orlistat, was accomplished with Docoviz pipeline (Fig 6). Docoviz is an automated system used for docking of drugs against protein structures at large scale using Auto-dock Vina [98]. This system is based upon perl and other languages such as ruby (manuscript in preparation). We obtained structural information of the genes implicated in obesity from protein data bank (PDB). Orlistat as well as other drugs were obtained from Drugbank [99] and their side effects were retrieved

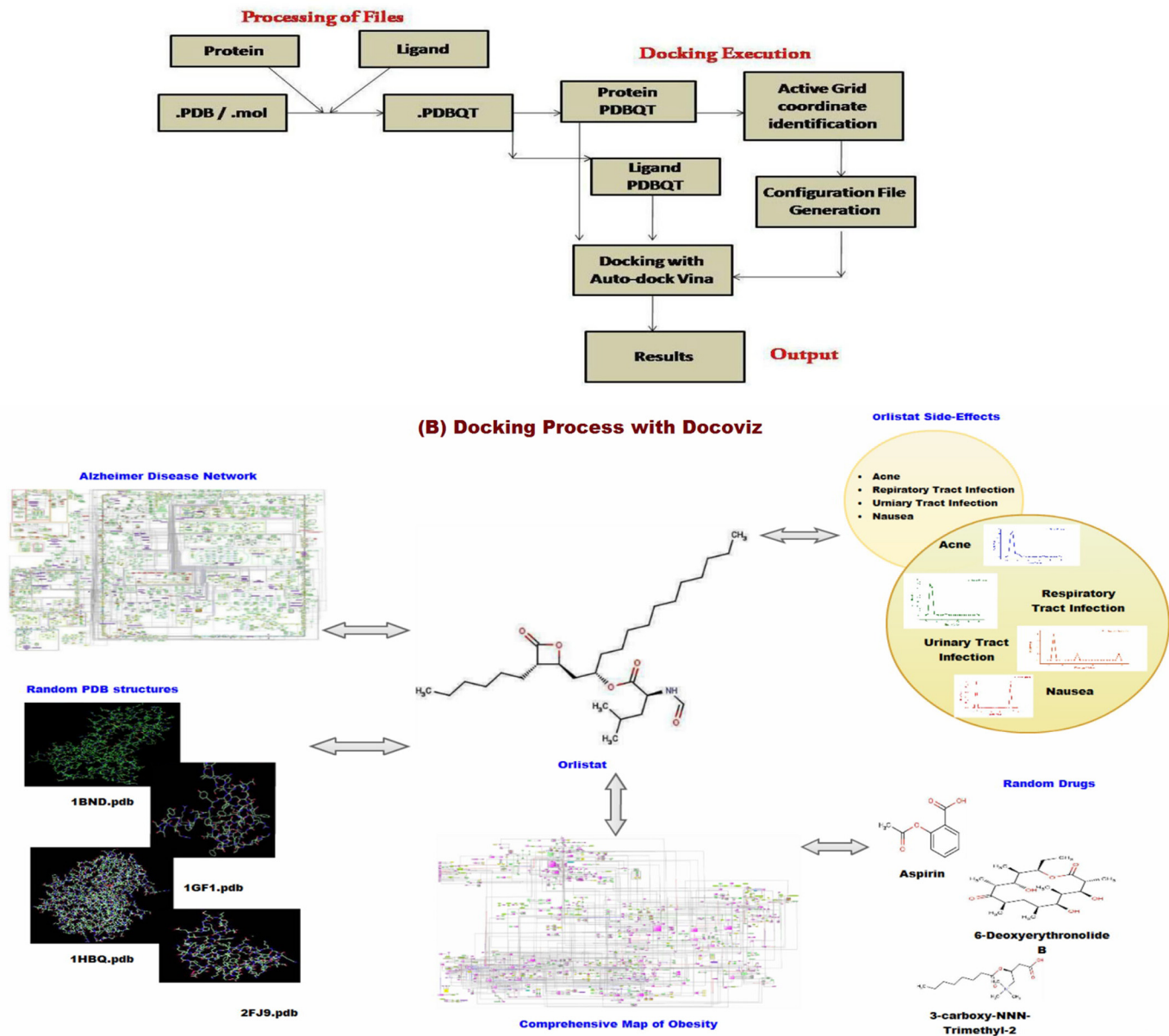


Fig 6. Shows the (A) schematic diagram of Docoviz pipeline and its (B) applications.

doi:10.1371/journal.pone.0146759.g006



from SIDER database. The pdb format of protein structure was converted to pdbqt format before commencing the docking procedure. We identified active site coordinates through geometric search method. A grid of about 20Å around the active site coordinates was generated to search all possible transition point (See K File in [S1 File](#)).

## Supporting Information

**S1 File.**  
(DOC)

**S2 File. Tables A through J are present in this document.**  
(XLS)

**S3 File. It contains information on websites containing additional supplementary data.**  
(DOCX)

## Acknowledgments

We thank Professor Ram Ramaswamy for editing this paper and for helpful suggestions. We thank Jaypee Institute of Information Technology, for their constant support. We are thankful to Indian Institute of Technology, New Delhi for providing the access to their super-computing facility for the execution of our programs. We also thank anonymous reviewers for their valuable comments.

## Author Contributions

Conceived and designed the experiments: KR. Performed the experiments: KR JJ HKJ SA. Analyzed the data: KR JJ HKJ. Contributed reagents/materials/analysis tools: KR SA HKJ. Wrote the paper: KR JJ HKJ.

## References

1. Pontzer H, Raichlen DA, Wood BM, Mabulla AZP, Racette SB. Energetics and Human Obesity. *Plos One*. 2012; 7: e40503. doi: [10.1371/journal.pone.0040503](https://doi.org/10.1371/journal.pone.0040503) PMID: [22848382](https://pubmed.ncbi.nlm.nih.gov/22848382/)
2. Mokdad AH, Serdula MK, Dietz WH, Bowman BA, Marks JS. The spread of the Obesity Epidemic in the United States, 1991–1998. *JAMA* 282: 1519–1522. PMID: [10546690](https://pubmed.ncbi.nlm.nih.gov/10546690/)
3. Mokdad AH, Ford ES, Bowman BA, Dietz WH, Vinicor F. Prevalence of Obesity, Diabetes, and Obesity-Related Health Risk Factors. *JAMA*.2003; 289: 176–179.
4. Turnbaugh PJ, Gordon JI. The core gut microbiome, energy balance and obesity. *J. Physiol.*2009; 587: 4153–4158. doi: [10.1113/jphysiol.2009.174136](https://doi.org/10.1113/jphysiol.2009.174136) PMID: [19491241](https://pubmed.ncbi.nlm.nih.gov/19491241/)
5. Verma A, Jayaraman M, Kumar HKVS, Modi KD. Hypertension and obesity causes or effect. *Saudi. Med. J.* 2008; 29:1135–1138. PMID: [18690306](https://pubmed.ncbi.nlm.nih.gov/18690306/)
6. Viner RM, Hsia Y, Tomsic T, Wong IC. Efficacy and safety of anti-obesity drugs in children and adolescents: systematic review and meta-analysis. *Obes. Rev.*2010; 11: 593–602. doi: [10.1111/j.1467-789X.2009.00651.x](https://doi.org/10.1111/j.1467-789X.2009.00651.x) PMID: [19922432](https://pubmed.ncbi.nlm.nih.gov/19922432/)
7. Elangbam CS. Review Paper: Current Strategies in the Development of Anti-obesity Drugs and Their Safety Concerns. *Vet. Pathol.* 2009; 46: 10–24. doi: [10.1354/vp.46-1-10](https://doi.org/10.1354/vp.46-1-10) PMID: [19112110](https://pubmed.ncbi.nlm.nih.gov/19112110/)
8. Mitka M. Surgery for Obesity. *JAMA*. 2003; 289: 1761–1762. PMID: [12684340](https://pubmed.ncbi.nlm.nih.gov/12684340/)
9. Miller WC, Kocaja DM, Hamilton EJ. A meta-analysis of the past 25 years of weight loss research using diet, exercise or diet plus exercise intervention. *Int J. Obes. Relat. Metab. Disord.* 1997; 10: 941–947.
10. Itariu BK, Zeyda M, Hochbrugger EE, Neuhofer A, Prager G, Schindler K, Bohdjalian A, Mascher D, Vangala S, Schranz M, Krebs M, Bischof MG, Stulnig TM. Longchain nPUFAs reduce adipose tissue and systemic inflammation in severely obese nondiabetic patients: a randomized controlled trial. *Am J Clin Nutr.* 2012; 96:1137–1149. doi: [10.3945/ajcn.112.037432](https://doi.org/10.3945/ajcn.112.037432) PMID: [23034965](https://pubmed.ncbi.nlm.nih.gov/23034965/)

11. Tseng YH, Kokkotou E, Schulz TJ, Huang TL, Winnay JN. New role of bone morphogenetic protein 7 in brown adipogenesis and energy expenditure. *Nature*. 2008; 454: 1000–1004. doi: [10.1038/nature07221](https://doi.org/10.1038/nature07221) PMID: [18719589](https://pubmed.ncbi.nlm.nih.gov/18719589/)
12. Gesta S, Tseng YH, Kahn CR. Developmental Origin of Fat: Tracking Obesity to its source. *Cell*. 2007; 131: 242–256. PMID: [17956727](https://pubmed.ncbi.nlm.nih.gov/17956727/)
13. Price RA, Stunkard AJ. Commingling analysis of obesity in twins. *Hum. Hered*. 1989; 39: 121–135. PMID: [2591975](https://pubmed.ncbi.nlm.nih.gov/2591975/)
14. Price RA, Ness R, Laskarzewski P. Common major gene inheritance of extreme overweight. *Hum. Biol*. 1990; 62: 747–765. PMID: [2262200](https://pubmed.ncbi.nlm.nih.gov/2262200/)
15. Orchard S, Kerrien S, Abbani S, Aranda B, Bhate J. Protein interaction data curation: the International Molecular Exchange [IMEx] consortium. *Nature Methods*. 2012; 9: 345–350. doi: [10.1038/nmeth.1931](https://doi.org/10.1038/nmeth.1931) PMID: [22453911](https://pubmed.ncbi.nlm.nih.gov/22453911/)
16. Tieri P, Fuente ADL, Termanini A, Franceschi C. Integrating Omics Data for Signaling Pathways, Interactome Reconstruction, and Functional Analysis. *Meth. Mol. Biol*. 2011; 719: 415–433.
17. Goh KI, Cusick ME, Valle D, Childs B, Vidal M, et al. The human disease network. *Proc Natl Acad Sci USA*. 2007; 104: 8685–8690. PMID: [17502601](https://pubmed.ncbi.nlm.nih.gov/17502601/)
18. Janjic V, Pržulj N. The core diseasome. *Molecular Biosystems*. 2012; 8: 2614–2625. doi: [10.1039/c2mb25230a](https://doi.org/10.1039/c2mb25230a) PMID: [22820726](https://pubmed.ncbi.nlm.nih.gov/22820726/)
19. Lusic AJ, Weiss JN. Cardiovascular networks: systems-based approaches to cardiovascular disease. *Circulation*. 2010; 121: 157–170. doi: [10.1161/CIRCULATIONAHA.108.847699](https://doi.org/10.1161/CIRCULATIONAHA.108.847699) PMID: [20048233](https://pubmed.ncbi.nlm.nih.gov/20048233/)
20. Wu G, Zhu L, Dent JE, Nardini C. A Comprehensive Molecular Interaction Map for Rheumatoid Arthritis. *PLoS One*. 2010; 5: e10137. doi: [10.1371/journal.pone.0010137](https://doi.org/10.1371/journal.pone.0010137) PMID: [20419126](https://pubmed.ncbi.nlm.nih.gov/20419126/)
21. Kohn KW. Molecular interaction map of the mammalian cell cycle control and DNA repair systems. *Mol. Biol. Cell*. 1999; 10: 2703–2734. PMID: [10436023](https://pubmed.ncbi.nlm.nih.gov/10436023/)
22. Oda K, Matsuoka Y, Funahashi A, Kitano H. A comprehensive pathway map of epidermal growth factor receptor signaling. *Mol. Syst. Biol*. 2005; 1: 1–17.
23. Oda K, Kitano H. A comprehensive map of the toll-like receptor signaling network. *Mol. Syst. Biol*. 2006; 2: 2006.0015. PMID: [16738560](https://pubmed.ncbi.nlm.nih.gov/16738560/)
24. Calzone L, Gelay A, Zinovyev A, Radvanyi F, Barillot E. A comprehensive modular map of molecular interactions in RB/E2F pathway. *Mol. Syst. Biol*. 2008; 4: 1–12.
25. Allison DB, Cui X, Page GP, Sabripour M. Microarray data analysis: from disarray to consolidation and consensus. *Nat. Rev. Genet*. 2006; 7: 55–56. PMID: [16369572](https://pubmed.ncbi.nlm.nih.gov/16369572/)
26. Schwikowski B, Uetz P, Fields S. A network of protein–protein interactions in yeast. *Nat. Biotech*. 2000; 18: 1257–1261.
27. Rual JF, Venkatesan K, Hao T, Hirozane-Kishikawa T, Dricot A. Towards a proteome-scale map of the human protein–protein interaction network. *Nature*. 2005; 437: 1173–1178. PMID: [16189514](https://pubmed.ncbi.nlm.nih.gov/16189514/)
28. Meyer M, Lorscheid I, Troitzsch KG. The Development of Social Simulation as Reflected in the First Ten Years of JASSS: a Citation and Co-Citation Analysis. *J. ASSS*. 2009; 12: 1–20.
29. Li S, Wu LJ, Zhang ZQ (2006) Constructing biological networks through combined literature mining and microarray analysis: a LMMA approach. *Bioinformatics* 22: 2143–2150. PMID: [16820422](https://pubmed.ncbi.nlm.nih.gov/16820422/)
30. Rankinen T, Zuberi A, Chagnon YC, Weisnagel SJ, Argyropoulos G. The human obesity genemap: the 2005 update. *Obesity [Silver Spring]*. 2006; 14: 529–644.
31. Bravard A, Lefai E, Meugnier E, Pesenti S, Disse E. FTO is increased in muscle during type 2 diabetes, and its overexpression in myotubes alters insulin signaling, enhances lipogenesis and ROS production, and induces mitochondrial dysfunction. *Diabetes*. 2011; 60: 258–268. doi: [10.2337/db10-0281](https://doi.org/10.2337/db10-0281) PMID: [20943749](https://pubmed.ncbi.nlm.nih.gov/20943749/)
32. Ichimura A, Hirasawa A, Poulain-Godefroy O, Bonnefond A, Hara T. Dysfunction of lipid sensor GPR120 leads to obesity in both mouse and human. *Nature*. 2012; 19: 350–354.
33. Hearn T, Spalluto C, Phillips VJ, Renforth GL, Copin N. Subcellular localization of ALMS1 supports involvement of centrosome and basal body dysfunction in the pathogenesis of obesity, insulin resistance, and type 2 diabetes. *Diabetes*. 2005; 54: 1581–1587. PMID: [15855349](https://pubmed.ncbi.nlm.nih.gov/15855349/)
34. Ebert MH, Schmidt DE, Thompson T, Butler MG. Elevated plasma gamma-aminobutyric acid [GABA] levels in individuals with either Prader-Willi syndrome or Angelman syndrome. *J. Neuropsychiatry Clin. Neurosci*. 1997; 9: 75–80. PMID: [9017532](https://pubmed.ncbi.nlm.nih.gov/9017532/)
35. Funahashi A, Tanimura N, Morohashi M, Kitano H. CellDesigner: a process diagram editor for gene-regulatory and biochemical networks. *Biosilico*. 2003; 1: 159–162.

36. Le Novère N, Hucka M, Mi H, Moodie S, Schreiber F. The systems biology graphical notation. *Nat Biotechnol.* 2009; 27: 735–741. doi: [10.1038/nbt.1558](https://doi.org/10.1038/nbt.1558) PMID: [19668183](https://pubmed.ncbi.nlm.nih.gov/19668183/)
37. Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* 2000; 8: 23–30.
38. Haw R, Croft D, Yung CK, et al. The Reactome BioMart. Database. 2011; 2011:1–7.
39. Hogart A, Nagarajan RP, Patzel KA, Yasui DH, LaSalle JM. 15q11-13 GABAA receptor genes are normally biallelically expressed in brain yet are subject to epigenetic dysregulation in autism-spectrum disorders. *Hum Mol Genet.* 2007; 16: 691–703. PMID: [17339270](https://pubmed.ncbi.nlm.nih.gov/17339270/)
40. Uruno A, Matsuda K, Noguchi N, Yoshikawa T, Kudo M. Peroxisome proliferator-activated receptor- $\alpha$  suppresses CYP11B2 expression and aldosterone production. *J Mol Endocrinology.* 2011; 46: 37–49.
41. Kitano H, Oda K, Kimura T, Matsuoka Y, Csete M. Metabolic Syndrome and Robustness Tradeoffs. *Diabetes.* 2004; 53: S6–S15. PMID: [15561923](https://pubmed.ncbi.nlm.nih.gov/15561923/)
42. Hartwell LH, Hopfield JJ, Leibler S, Murray AW. From molecular to modular cell biology. *Nature.* 1999; 402: C47–C52. PMID: [10591225](https://pubmed.ncbi.nlm.nih.gov/10591225/)
43. Higham D, Kalna G, Kibble M. Spectral clustering and its use in bioinformatics. *J. Comp. Appli. Math.* 2007; 204: 25–37.
44. Newman MEJ. Finding community structure in networks using the eigenvectors of matrices. *Phys. Rev. E. Stat. Nonlin. Soft Matter Phys.* 2006; 74: 1–22.
45. Karger D, Stein C. A new approach to the minimum cut problem. *J. ACM.* 1996; 43: 601–640.
46. Brandes U, Delling D, Gaertler M, Orke RG, Hofer M. On Finding Graph Clustering with Maximum Modularity. *Graph-Theoretic Concepts in Computer Science.* 2007; 121–132.
47. Chung F. The heat kernel as the page rank of a graph. *Proc. Natl. Acad. Sci. USA.* 2007; 104: 19735–19740.
48. Yoon J, Blumer A, Lee K. An algorithm for modularity analysis of directed and weighted biological networks based on edge-betweenness centrality. *Bioinformatics.* 2006; 22: 3106–3108. PMID: [17060356](https://pubmed.ncbi.nlm.nih.gov/17060356/)
49. Ucar D, Asur S, Catalyurek U, Parthasarathy S. Improving functional modularity in protein-protein interactions graphs using hub-induced subgraphs. *Lecture Notes in Computer Science.* 2006; 4213: 371–371.
50. Bader GD, Hogue CW. An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics.* 2003; 4: 1–27.
51. Spirin V, Mirny LA. Protein complexes and functional modules in molecular networks. *Proc. Natl. Acad. Sci. USA.* 2003; 100: 12123–12128. PMID: [14517352](https://pubmed.ncbi.nlm.nih.gov/14517352/)
52. Dhillon IS, Guan Y, Kulis B. Weighted graph cuts without eigenvectors: A multilevel approach. *IEEE Trans. Pattern. Anal. Mach. Intell.* 2007; 29: 1944–1957. PMID: [17848776](https://pubmed.ncbi.nlm.nih.gov/17848776/)
53. Newman NEJ. Modularity and community structure in networks. *Proc. Natl. Acad. Sci. USA.* 2006; 103:8577–8582. PMID: [16723398](https://pubmed.ncbi.nlm.nih.gov/16723398/)
54. Fruhbeck G. Intracellular signalling pathways activated by leptin. *Biochem J.* 2006; 393: 7–20. PMID: [16336196](https://pubmed.ncbi.nlm.nih.gov/16336196/)
55. Tam CS, Lecoultré V, Ravussin E. Novel strategy for the use of leptin for obesity therapy. *Expert. Opin. Biol. Ther.* 2011; 11: 1677–1685. doi: [10.1517/14712598.2011.619974](https://doi.org/10.1517/14712598.2011.619974) PMID: [21910668](https://pubmed.ncbi.nlm.nih.gov/21910668/)
56. Yamada N, Katsuura G, Ochi Y, Ebihara K, Kusakabe T. Impaired CNS Leptin Action Is Implicated in Depression Associated with Obesity. *Endocrinology.* 2003; 152: 2634–2643.
57. Toshinai K, Date Y, Murakami N, Shimada M, Mondal MS. Ghrelin-Induced Food Intake Is Mediated via the Orexin Pathway. *Endocrinology.* 2003; 144: 1506–1512. PMID: [12639935](https://pubmed.ncbi.nlm.nih.gov/12639935/)
58. Nogueiras R, Tschöp MH, Zigman JM. Central nervous system regulation of energy metabolism: ghrelin versus leptin. *Ann. N. Y. Acad. Sci.* 2008; 1126: 14–19. doi: [10.1196/annals.1433.054](https://doi.org/10.1196/annals.1433.054) PMID: [18448790](https://pubmed.ncbi.nlm.nih.gov/18448790/)
59. Scerif M, Goldstone AP, Korbonits M. Ghrelin in obesity and endocrine diseases. *Mol. Cell. Endocrinol.* 2011; 340: 15–25. doi: [10.1016/j.mce.2011.02.011](https://doi.org/10.1016/j.mce.2011.02.011) PMID: [21345363](https://pubmed.ncbi.nlm.nih.gov/21345363/)
60. Zhang N, Yuan C, Li Z, Li J, Li X. Meta-Analysis of the Relationship Between Obestatin and Ghrelin Levels and the Ghrelin/Obestatin Ratio With Respect to Obesity. *Am J Med Sci.* 2011; 341: 48–55. doi: [10.1097/MAJ.0b013e3181ec41ed](https://doi.org/10.1097/MAJ.0b013e3181ec41ed) PMID: [21139496](https://pubmed.ncbi.nlm.nih.gov/21139496/)
61. Niswender KD, Daws LC, Avison MJ, Galli A. Insulin Regulation of Monoamine Signaling: Pathway to Obesity. *Neuropsychopharmacology.* 2011; 36: 359–360.
62. Gallagher EJ, LeRoith D. Insulin, insulin resistance, obesity, and cancer. *Curr. Diab. Rep.* 2010; 10: 93–100. doi: [10.1007/s11892-010-0101-y](https://doi.org/10.1007/s11892-010-0101-y) PMID: [20425567](https://pubmed.ncbi.nlm.nih.gov/20425567/)

63. Toh SA, Levin M, Rader DJ. Atherogenic Lipid Metabolism in Obesity. *Met. basis obes.* 2011; 293–309.
64. Purnell JQ., Hokanson JE, Cleary PA, Nathan DM, Lachin JM. The Effect of Excess Weight Gain with Intensive Diabetes Treatment on Cardiovascular Disease Risk Factors and Atherosclerosis in Type 1 Diabetes: Results from the Diabetes Control and Complications Trial / Epidemiology of Diabetes Interventions and Complications Study [DCCT/EDIC] Study. *Circulation.* 2012; 127: 180–187. doi: [10.1161/CIRCULATIONAHA.111.077487](https://doi.org/10.1161/CIRCULATIONAHA.111.077487) PMID: [23212717](https://pubmed.ncbi.nlm.nih.gov/23212717/)
65. Prestwich TC, Macdougald OA. Wnt/ $\beta$ -catenin signaling in adipogenesis and metabolism. *Curr. Opin. Cell. Biol.* 2007; 19: 612–617. PMID: [17997088](https://pubmed.ncbi.nlm.nih.gov/17997088/)
66. Lee H, Bae S, Yoon Y. WNT/ $\beta$ -catenin pathway mediates the anti-adipogenic mechanism of SH21B, a traditional herbal medicine for the treatment of obesity. *J. Ethnopharmacol.* 2011; 133: 788–795. doi: [10.1016/j.jep.2010.11.009](https://doi.org/10.1016/j.jep.2010.11.009) PMID: [21070846](https://pubmed.ncbi.nlm.nih.gov/21070846/)
67. Ravasz E, Somera AL, Mongru DA, Oltvai ZN, Barabasi AL. Hierarchical organization of modularity in metabolic networks. *Science.* 2002; 297: 1551–1555. PMID: [12202830](https://pubmed.ncbi.nlm.nih.gov/12202830/)
68. DeSilva E, Stumpf MPH. Complex networks and simple models in biology. *J R Soc Interface.* 2005; 2: 419–430. PMID: [16849202](https://pubmed.ncbi.nlm.nih.gov/16849202/)
69. Barrat A, Barthélemy M, Vespignani A. Modeling the evolution of weighted networks. *Phys Rev E Stat Nonlin Soft Matter Phys.* 2004; 70: 066149. PMID: [15697476](https://pubmed.ncbi.nlm.nih.gov/15697476/)
70. Bourguignon P-Y, Danos V, Kepes FC, Smidtas S, Schachter V. Property-Driven Statistics of Biological Networks. *Trans on Comput Syst Biol VI.* 2006; 4220: 1–15.
71. Erdos P, Renyi A. On the Evolution of Random Graphs. *Publ. Math. Inst. Hung. Acad. Sci.* 1960; 5: 17–61.
72. Watts DJ, Strogatz SH. Collective dynamics of ‘small-world’ networks. *Nature.* 1998; 393: 440–442. PMID: [9623998](https://pubmed.ncbi.nlm.nih.gov/9623998/)
73. Barabasi AL, Albert R. Emergence of scaling in random networks. *Science.* 1999; 286:509–12. PMID: [10521342](https://pubmed.ncbi.nlm.nih.gov/10521342/)
74. Maere S, Heymans K, Kuiper M. BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics.* 2005; 15:3448–3449.
75. Scardoni G, Laudanna C. Network centralities Interference and Robustness. *Int. J. Complex Systems in Science.* 2011; 164–168.
76. Cohen R., Erez K., Ben-Avraham D. & Havlin S. Resilience of the Internet to random breakdowns. *Phys. Rev. Lett.* 2000; 85:4626–4628. PMID: [11082612](https://pubmed.ncbi.nlm.nih.gov/11082612/)
77. Pastor-Satorras R. & Vespignani A. Epidemic spreading in scale-free networks. *Phys. Rev. Lett.* 2001; 86: 3200–3203. PMID: [11290142](https://pubmed.ncbi.nlm.nih.gov/11290142/)
78. Nishikawa T., Motter A. E., Lai Y.-C. & Hoppensteadt F. C. Heterogeneity in oscillator networks: are smaller worlds easier to synchronize? *Phys. Rev. Lett.* 2003; 91:014101. PMID: [12906539](https://pubmed.ncbi.nlm.nih.gov/12906539/)
79. Wang J, Huang Q, Liu ZP, Wang Y, Wu LY. NOA: a novel NetworkOntologyAnalysis method. *Nucleic Acids Res.* 2011; 39:e87. doi: [10.1093/nar/gkr251](https://doi.org/10.1093/nar/gkr251) PMID: [21543451](https://pubmed.ncbi.nlm.nih.gov/21543451/)
80. Kapushesky M, Adamusiak T, Burdett T, Culhane A, Farné A. Gene Expression Atlas update—a value-added database of microarray and sequencing-based functional genomics experiments. *Nucleic Acids Res.* 2012; D1077–81. doi: [10.1093/nar/gkr913](https://doi.org/10.1093/nar/gkr913) PMID: [22064864](https://pubmed.ncbi.nlm.nih.gov/22064864/)
81. Mizuno S, Iijima R, Ogishima S, Kikuchi M, Matsuoka Y. AlzPathway: a comprehensive map of signaling pathways of Alzheimer’s disease. *BMC Syst Biol.* 2012; 6:52. doi: [10.1186/1752-0509-6-52](https://doi.org/10.1186/1752-0509-6-52) PMID: [22647208](https://pubmed.ncbi.nlm.nih.gov/22647208/)
82. Hopkins AL. Networkpharmacology: the next paradigm in drug discovery. *Nat Chem Biol.* 2008; 4:682–90. doi: [10.1038/nchembio.118](https://doi.org/10.1038/nchembio.118) PMID: [18936753](https://pubmed.ncbi.nlm.nih.gov/18936753/)
83. Gao Z, Li H, Zhang H, Liu X, Kang L. PDTD: a web-accessible protein database for drug target identification. *BMC bioinformatics.* 2008; 9:104. doi: [10.1186/1471-2105-9-104](https://doi.org/10.1186/1471-2105-9-104) PMID: [18282303](https://pubmed.ncbi.nlm.nih.gov/18282303/)
84. Hui-fang L, Qing S, Jian Z, Wei F. Evaluation of various inverse docking schemes in multiple targets identification. *Journal of molecular graphics & modelling.* 2010; 29: 326–330.
85. Lee M, Kim D. Large-scale reverse docking profiles and their applications. *BMC Bioinformatics.* 2012; 13:S6.
86. Smith SC, Havaleshko DM, Moon K, Baras AS, Lee J. Use of yeast chemigenomics and COXEN informatics in preclinical evaluation of anticancer agents. *Neoplasia.* 2011; 13:72–80. PMID: [21253455](https://pubmed.ncbi.nlm.nih.gov/21253455/)
87. Jeong H, Mason SP, Barabasi AL, Oltvai ZN. Lethality and centrality in protein networks. *Nature.* 2001; 411: 41–42. PMID: [11333967](https://pubmed.ncbi.nlm.nih.gov/11333967/)
88. Fraser HB, Hirsh AE, Steinmetz LM, Scharfe C, Feldman MW. Evolutionary rate in the protein interaction network. *Science.* 2002; 296: 750–752.

89. Eisenberg E, Levanon EY. Preferential attachment in the protein network evolution. *Phys. Rev. Lett.* 2003; 91: 138701. PMID: [14525344](#)
90. Saeed R, Deane CM. Protein protein interactions, evolutionary rate, abundance and age. *BMC Bioinformatics.* 2006; 7: 1–13.
91. Yu H, Braun P, Yildirim MA, Lemmens I, Venkatesan K. High-quality binary protein interaction map of the yeast interactome network. *Science.* 2008; 322: 104–110. doi: [10.1126/science.1158684](#) PMID: [18719252](#)
92. Jordan IK, Wolf YI, Koonin EV. No simple dependence between protein evolution rate and the number of protein-protein interactions: only the most prolific interactors tend to evolve slowly. *BMC Evol. Biol.* 2003; 3:1–8. PMID: [12515583](#)
93. Logsdon BA, Hoffman GE, Mezey JG. Mouse obesity network reconstruction with a variational Bayes algorithm to employ aggressive false positive control. *BMC Bioinformatics.* 2012; 13: 1–15.
94. Brown CD, Higgins M, Donato KA, Rohde FC, Garrison R. Body Mass Index and the Prevalence of Hypertension and Dyslipidemia. *Obesity Research.* 2012; 8: 605–619.
95. Aronne LJ. Classification of obesity and assessment of obesity-related health risks. *Obes Res.* 2002; 10 Suppl 2:105S–115S. PMID: [12490659](#)
96. Brohee S, Faust K, Lima-Mendez GSO, Janky R, Vanderstocken G. NeAT: a toolbox for the analysis of biological networks, clusters, classes and pathways. *Nucleic Acids Res.* 2008; 38: W444–W451.
97. Assenov Y, Ramírez F, Schelhorn SE, Lengauer T, Albrecht M. Computing topological parameters of biological networks. *Bioinformatics.* 2008; 24: 282–284. PMID: [18006545](#)
98. Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput, Chem.* 2010; 31: 455–61.
99. Knox C, Law V, Jewison T, Liu P, Ly S. DrugBank 3.0: a comprehensive resource for 'omics' research on drugs. *Nucleic Acids Res.* 2011; 39: D1035–41. doi: [10.1093/nar/gkq1126](#) PMID: [21059682](#)