We are IntechOpen, the world’s leading publisher of Open Access books
Built by scientists, for scientists

4,300
Open access books available

116,000
International authors and editors

130M
Downloads

154
Countries delivered to

TOP 1%
Our authors are among the most cited scientists

12.2%
Contributors from top 500 universities

WEB OF SCIENCE™
Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com
Network Analysis of Obesity Expression Data

Ashok Kumar and Kanika Gupta

Abstract

There are numerous genetic and environmental factors associated significantly with obesity, which could be used as potential diagnostic biomarkers. The molecular mechanisms, development, differentiation, and disease gene expression data provide crucial insights as these differentially expressed genes could have major effects on diet-induced obesity and such effect is not seen in animals. Genomics and proteomics are major branches for better understanding the normal function of the tissues and their interactions with the environment i.e. characterizing the tissues in which the newly discovered genes are expressed, helps in understanding the development of tissues, ageing mechanisms, and signalling routes that enable the tissues to function and also direct the similitude, parallelism and other levels of aptness between two or more gene artefacts. It is traditionally known that hypothalamic and brain stem centres are intricate in the mandate of food absorption and energy equilibrium, but statistics on the associated governing elements and their genes was scant until the utmost decagon and have been identified to be strongly expressed in variety of tissues. NPY plays a notable part in anxiety, tension, corpulence, and vitality homeostasis through incitement of NPY-Y1 receptors (Y1Rs) in the mind. NPY1R quality is the protein accomplice of qualities that are utilized as model as a part of mouse and in addition in people. Utilizing diverse bioinformatics instruments, the relative examination of NPY1R at quality and additionally at protein level can be assessed for biomarker of stoutness malady. In this manner, the system science thinks about point to predict the quality of heftiness which could be taken as a biomarker in human by examining with the quality that already has been utilized as marker as a part of model life forms.

Keywords: network biology, text mining, obesity, bioinformatics databases and tools
1. Introduction

Creation of networks and all their known associations [1], enabled valuable insights into human disease and disease therapy. Protein-protein interaction mapping focused on specific human diseases which identified novel interactions among proteins encoded by known disease genes, and have also predicted new disease susceptibility genes. Rapid advances in network biology indicated that cellular networks are governed by universal laws and offer a new conceptual framework that could potentially revolutionize our view of biology and disease pathologies in the twenty-first century [2]. Due to the wide quota of research being conducted on this topic, much has been inscribed in the biomedical literature about the coalition betwixt genes and diseases. Therefore, obtaining disease–gene coalition from script is an evident use case for text mining, and disease–gene coalitions have actually formerly been obtained by postulated co-occurrence-based text-mining structures [3–6]. Text mining is the discovery by computer of new, previously unknown information, by automatically extracting information from different written resources. The purpose of text mining is to process unstructured (textual) information, extract meaningful numeric indices from the text, and, thus, make the information contained in the text accessible to the various data mining (statistical and machine learning) algorithms. As the research on obesity is carried out by large groups in scientific community, this becomes the problem of big data analytics that is, the process of examining large data sets containing a variety of data types to uncover hidden patterns and unknown correlations. Obesity is an abnormal accumulation of body fat, usually 20% or more over an individual's ideal body weight. Excess bodyweight is the sixth most important risk factor contributing to the overall burden of disease worldwide. Genetic factors significantly influence how the body regulates the appetite and the rate at which it turns food into energy (metabolic rate). A lot is known about the genetic aspects of obesity, but much more remains to be discovered. The primary goals are to identify the specific genetic variations and the biologic consequences that are produced, or as commonly put, discovering the genes and pathways involved in producing phenotypic variation and the factors that influence obesity [7]. Thus from the present work we would find markers for obesity in humans which would help in the diagnosis and prognosis of obesity and the same process could find its applications for other diseases.

2. Network biology and text mining approach to find potential human biomarkers in obesity

Network science concerns with biological entanglement by condensing composite structures as elements (nodes) and interactions (edges) betwixt them [8]. In biological structures nodes are metabolites and macromolecules such as proteins, RNA molecules and gene sequences, while the edges are physical, biochemical and functional interactions that can be recognised with a profusion of automation. Creation of networks of genetic disorders and all their known gene associations [1], or of drugs and all their known protein targets [9], enabled worthwhile insights into human disease and disease therapy. Protein-protein interaction mapping efforts
focused on specific human diseases (like ataxia [10, 11], autism [12] and breast cancer [13] have identified novel interactions among proteins encoded by known disease genes, and have also predicted new disease susceptibility genes. The common finding among these disease interactomes is the discovery of unexpected relationships between disease genes that initially appeared unrelated [14]. Building and analysing more disease-centric networks is accordingly a critical step towards deeper understanding of underlying disease mechanisms (http://ccsb.dfci.harvard.edu/web/www/ccsb/Research/networks.html). A key aim of postgenomic biomedical research is to systematically catalogue all molecules and their interactions within a living cell as shown in Figure 1. There is a comprehensible necessity to comprehend how these molecules and the interactions between them decide the role of this extremely composite mechanism, both in detachment and when encompassed by different cells. Fast advances in system science determine that cell systems are hegemonize by general laws and offer another calculated structure that could change the perspective of science and infection pathologies in the twenty-first century [2].

Figure 1. Uproars in Biological Systems and Cellular Networks may stamp genotype-phenotype connections. By communicating with each other, qualities and their items from complex cell systems. The connection between upheavals in system and frameworks properties and phenotypes, for example, Mendelian issue, complex qualities, and tumour, may be as major as that amongst genotypes and phenotypes [8].

Three distinct approaches have been used to capture interactome networks: (1) compilation or curation of hitherto prevailing data accessible in the writing, more often than not removed from one or only a couple sorts of physical or biochemical associations [15]; (2) computational expectations in light of available “orthogonal” data separated from physical or biochemical collaborations, for example, arrangement likenesses, quality request protection, co-nearness and co-nonappearance of qualities in totally sequenced genomes and protein basic data [16]; and (3) orderly, unprejudiced high throughput experimental mapping strategies applied at the scale of whole genomes or proteomes [17]. These approaches, though compatible, differ greatly
in the feasible interpretations of the resulting maps. Literature-curated maps extend the benefit of using already accessible information, but are restricted by the intrinsically variable quality of the published data, the absence of orderliness, and the absence of describing of negative data [18, 19]. Computational prediction maps are fast and efficient to implement, and usually include satisfyingly large numbers of nodes and edges, but are necessarily imperfect because they use indirect information [20]. While high-throughput maps attempt to report unbiased, deliberate, and all around controlled information, they were at first all the more difficult to start, albeit late mechanical methodology predict that close achievement can come within a couple of years for profoundly reliable, comprehensive protein-protein connection and quality administrative system maps for human [21]. Content mining is the disclosure by PC of new, beforehand obscure data, by normally acquiring data from various composed courtesy. A key part is the association of the acquired data together to frame new truths or new theories to be viewed as further by a more basic method for examination (http://people.ischool.berkeley.edu/~hearst/text-mining.html). The reason of text mining is to handle unstructured (literary) data, extricate important numeric records from the content, and, in this way, make the data required in the content accessible to the different information mining (factual and machine learning) techniques as shown in Figure 2. Data can be acquired to get synopses for the words required in the records or to register outlines for the archives in light of the words contained in them (http://documents.software.dell.com/statistics/textbook/text-mining# overview).

Figure 2. The approach followed by text mining method in general.

The heterogeneous data types are generated by experiments done. To communicate with these scientific discoveries natural language is used which is amenable for direct human interpretations. Natural language is the simple human language, different from programming lan-
guage, through which human talks to computer. Functional information and annotations can be derived from published text directly or indirectly. Currently databases are only capable of covering a small fraction of biological context information encountered in the literature. For bench scientists, published data is the best source for interpreting high-throughput experiments, but automated text processing methods are required to integrate them into the data analysis workflow. So, the user demands better information access that is beyond just keyword searches. Moreover, due to rapid growth of information, manual extraction of information is a difficult task. So, there is a need of an efficient approach that can retrieve the meaningful information from this vast and unstructured text [22]. Excess body weight is the sixth most important risk factor contributing to the overall burden of disease worldwide; 1.1 billion adults and 10% of children are now classified as overweight or obese. The main adverse consequences of being obese are cardiovascular disease, type 2 diabetes, and several cancers as shown in Figure 3 [23]. The incidence of obesity appears to be levelling in the world and started to be a big concern in the public health that causes social and economic costs of the twenty-first century. The pathogenesis of obesity is complex at all levels of biology as shown in Figure 4 that is genetics, cell and tissue biology, physiology, and behaviour. The International Diabetes Federation considers central obesity as a primary evidence of metabolic syndrome, with the additional features which include, (1) increased triglyceride levels, (2) increased blood pressure, (3) increased fasting plasma glucose and (4) reduced HDL-cholesterol [24]. In 1997, there was serious buoyancy because, for the first time in 25 years, a new drug for the treatment of obesity had been endorsed by the US Food and Drug Administration (FDA). Then, in April 1996, two more drugs were starting their way through the acceptance procedure [25, 26]. In June 2013, the American Medical Association classified obesity as a disease (http://www.medscape.com/viewarticle/806566).
A lot is known about the genetic aspects of obesity, but much more remains to be discovered. Medical genetics is fundamentally interested in understanding the relationship between genetic variation and human health and disease. The primary goals are to identify the specific genetic variations and the biologic consequences that are produced, or as commonly put, discovering the genes and pathways involved in producing phenotypic variation, and the factors that influence obesity [7]. Network study on genes and proteins offers functional basics of the complexity of gene and protein, and its interacting partners as shown in Figure 5. Obese adults and children are more likely to display elevations in plasma $\text{fabp4}$ levels [27, 28]. $\text{Ppar}\gamma$ appeared to be a core obesity gene, which interacts with lipid metabolism and inflammation genes [25]. Genetic variants within $\text{FTO}$ (fat mass and obesity associated) have been identified to exhibit the strongest association with obesity in humans [29–32]. The well-known obesity-related $\text{FTO}$ gene interacts with $\text{APOE}$ which in turn, is associated with Alzheimer’s disease [33] and with $\text{MC4R}$, resulting in a higher chance of breast cancer [34]. Gene networks can be constructed by ensembling previously reported interactions in the literature and various databases like STRING, DISEASES, etc. [35]. The network could be visualized and constructed using cytoscape. Cytoscape supported several algorithms for the layout of networks which included spring embedded layout, hierarchical layout, circular layout and attribute based layout [36]. It was generally accepted that hypothalamic and brain stem centres are involved in the regulation of food intake and energy balance but information on the relevant regulatory factors and their genes was scarce until the last decade [37].
There are numerous genetic factors, like Melanocortin-4 receptor (MC4R), Proopiomelanocortin (POMC), Single Minded Gene (SIM1), etc., important in obesity, which can be used as biomarkers in humans [38]. In the past literature studies, NPY1R was used as a knockout marker in mouse for obesity but not used as a biomarker in humans [39]. NPY1R (Neuropeptide Y Receptor Y1), have been recognized to actively express in variety of tissues, including trigeminal V ganglion, heart, brain, spleen, lungs, skeletal muscle, kidney and embryo, in embryonic as well as in postnatal Theiler stages as adamanted by RNA in situ and Northern blot [38, 40]. Therefore, interacting patterns of NPY1R were analysed using STRING version 10.0 [41] as shown in Figure 6.

Figure 5. Genes involved in the leptin-melanocortin pathway that have been associated with monogenic obesity through their influence on food intake and energy expenditure.

Figure 6. The interacting patterns of NPY1R in Homo sapiens obtained from known (curated databases and experimentally determined), predicted (gene-neighbourhood, gene fusions and gene co-occurrence) and other (text mining, protein homology and co-expression) interactions.
As NPY1R was used as an obesity marker in obesity model organisms like mouse and rat, therefore their interactions were also observed using STRING version 10.0 as shown in Figures 7 and 8.

Figure 7. The interacting patterns of NPY1R in Mus musculus obtained from known (curated databases and experimentally determined), predicted (gene-neighbourhood, gene fusions and gene co-occurrence) and other (text mining, protein homology and co-expression) interactions.

Figure 8. The interacting patterns of NPY1R in Rattus norvegicus obtained from known (curated databases and experimentally determined), predicted (gene-neighbourhood, gene fusions and gene co-occurrence) and other (text mining, protein homology and co-expression) interactions.

<table>
<thead>
<tr>
<th>Homo sapiens</th>
<th>Score</th>
<th>Mus musculus</th>
<th>Score</th>
<th>Rattus norvegicus</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>NPY</td>
<td>0.998</td>
<td>Npy</td>
<td>0.992</td>
<td>Npy</td>
<td>0.993</td>
</tr>
<tr>
<td>PYY</td>
<td>0.993</td>
<td>Gal</td>
<td>0.974</td>
<td>Gal</td>
<td>0.973</td>
</tr>
<tr>
<td>GAL</td>
<td>0.992</td>
<td>Pyy</td>
<td>0.965</td>
<td>Pyy</td>
<td>0.964</td>
</tr>
<tr>
<td>ppy</td>
<td>0.964</td>
<td>pmch</td>
<td>0.946</td>
<td>ppy</td>
<td>0.941</td>
</tr>
</tbody>
</table>

Table 1. Lists the top scoring functional partners in Homo sapiens, Mus musculus and Rattus norvegicus.
After finding the functional partners for NPY1R in human and obesity model organisms that is, mouse and rat, top four high scoring genes were considered and further their functional partners were retrieved from STRING version 10.0 as shown in Table 1. The score of the functional partners were mostly on the basis of known experimental and curated databases interactions, other interactions like text mining interactions.

The networks obtained from STRING for all the interactions were merged separately for three organisms using cytoscape version 2.7.0 as shown in Figures 9–11.

Figure 9. The merged network for Homo sapiens interacting functional partners. The green colour node shows the main input NPY1R for which the functional partners were searched. The sea green nodes show the top scoring functional partners of NPY1R.

Figure 10. The merged network for Mus musculus interacting functional partners. The green colour node shows the main input NPY1R for which the functional partners were searched. The sea green nodes show the top scoring functional partners of NPY1R.
Then these merged networks were manually analysed and it was found that there are 11 genes which were common in the merged networks of the three considered organisms. The common genes were *npy*, *ppy*, *pdyn*, *gal*, *pomc*, *npy1r*, *sst*, *galr1*, *npy2r*, *ccl28* and *npy5r*. Then these common genes were used to find disease-gene associations, in this case, association of common genes with obesity using DISEASES web source [42] that integrates evidence on disease-gene associations from automatic text mining, manually curated literature, cancer mutation data, and genome-wide association studies was found. From DISEASES web source 8 genes out of 11 were found related to obesity, where 7 genes had evidence from text mining and 1 gene had database evidence and no gene was found from experimental results as shown in Table 2.

<table>
<thead>
<tr>
<th>Gene Name</th>
<th>Disease</th>
<th>Evidence</th>
<th>Confidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>NPY</td>
<td>Obesity</td>
<td>Text mining</td>
<td>****</td>
</tr>
<tr>
<td>NPY1R</td>
<td>Obesity</td>
<td>Text mining</td>
<td>**</td>
</tr>
<tr>
<td>NPY2R</td>
<td>Obesity</td>
<td>Text mining</td>
<td>**</td>
</tr>
<tr>
<td>NPY5R</td>
<td>Obesity</td>
<td>Text mining</td>
<td>**</td>
</tr>
<tr>
<td>PYY</td>
<td>Obesity</td>
<td>Text mining</td>
<td>****</td>
</tr>
<tr>
<td>GAL</td>
<td>Obesity</td>
<td>Text mining</td>
<td>**</td>
</tr>
<tr>
<td>CCL28</td>
<td>Obesity</td>
<td>Text mining</td>
<td>**</td>
</tr>
<tr>
<td>POMC</td>
<td>Obesity</td>
<td>Database</td>
<td>****</td>
</tr>
</tbody>
</table>

*Table 2.* List of disease-genes associations acquired from automatic text mining of the biomedical literature and DISEASES web source, where the confidence of each association is signified by stars, where **** is the highest confidence and * is the lowest.
All the above gathered data was cross checked for networks and its disease associations using KEGG pathway [43, 44] which is a collection of manually drawn pathway maps representing the knowledge on the molecular interaction and reaction networks and Online Mendelian Inheritance in Man (OMIM) [45] which is a comprehensive, authoritative compendium of human genes and genetic phenotypes. Two pathways were found in humans which showed roles in obesity containing the respective genes obtained after disease-gene associations as shown in Figures 12 and 13.

![Figure 12](image1.png)

**Figure 12.** Regulation of lipolysis in adipocytes. This pathway shows the presence of genes NPYR and NPY in the fed state. This pathway also shows the presence of genes like FABP but in the fasting state and is the known marker for obesity [46–60].

Thus, from the above work we could conclude that NPY, NPY1R, NPY2R, NPY5R and POMC which in the past literature studies were used as knockout markers in mouse and rats for
obesity but not used as a biomarker in humans could be considered as potential biomarkers for obesity in humans. By finding optimal biomarkers, diagnostic criteria for cardiovascular diseases can be refined in the obese beyond “traditional” risk factors to identify early pathologic processes. Identifying diagnosis and prognosis biomarkers from expression profiling data is of great significance for achieving personalized medicine and designing a therapeutic strategy in complex diseases. A similar methodology can be used to predict other biomarkers for different diseases. For progression and maintenance of life saving diseases, the expression data of biomarkers could be used in future applications.

Figure 13. Adipocytokine signalling pathway. This pathway again marks the presence of NPY and POMC in obesity along with already known markers of obesity like PPAR and TNF α [61–74].

Author details

Ashok Kumar* and Kanika Gupta

*Address all correspondence to: ashokkumar@pu.ac.in

Centre For Systems Biology & Bioinformatics, UIEAST, Panjab University, Chandigarh
References


