Towards finding the linkage between metabolic and age-related disorders using semantic gene data network analysis

Mohammad Uzzal Hossain¹, Abu Zaffar Shibly¹, Taimur Md. Omar¹, Fatama Tous Zohora¹, Umme Sara Santona¹, Md. Jakir Hossain¹, Md. Sadek Hosen Khoka¹, Chaman Ara Keya², Md. Salimullah³

Received January 04, 2016; Accepted January 07, 2016; Published January 31, 2016

¹Department of Biotechnology and Genetic Engineering, Life Science Faculty, Mawlana Bhashani Science and Technology University, Santosh, Tangail 1902, Bangladesh; ḞDepartment of Biology and Chemistry, North south University, Bashundhara, Dhaka 1229, Bangladesh; ³Molecular Biotechnology Division, National Institute of Biotechnology, Ganakbari, Ashulia, Savar, Dhaka 1349, Bangladesh; Md. Salimullah (Director General – Additional Charge) – E-mail: salim2969@gmail.com; Phone: 880-2-7788443; Fax: 880-2-7789656; ḞCorresponding author

Abstract:
A metabolic disorder (MD) occurs when the metabolic process is disturbed. This process is carried out by thousands of enzymes participating in numerous inter-dependent metabolic pathways. Critical biochemical reactions that involve the processing and transportation of carbohydrates, proteins and lipids are affected in metabolic diseases. Therefore, it is of interest to identify the common pathways of metabolic disorders by building protein-protein interactions (PPI) for network analysis. The molecular network linkages between MD and age related diseases (ARD) are intriguing. Hence, we created networks of protein-associated proteins and predicted genes and or its products of ARD in common pathways. The genes in the common pathways were isolated from the network and further analyzed for their co-localization and shared domains. Thus, a model hypothesis is proposed using interaction networks that are linked between MD and ARD. This data even if less conclusive finds application in understanding the molecular mechanism of known diseases in relation to observed molecular events.

Keywords: Gene network, common pathways, metabolic disorders, age-related disorders, interactions

Background:
Metabolic disorder (MD) is a cluster of metabolic risk factors characterized by obesity, elevated blood pressure, increased plasma glucose (fasting), high triglycerides in serum and decreased high-density cholesterol levels [1]. Metabolic disorder affected people are at increased risk for atherosclerosis, peripheral vascular disease, coronary heart disease, myocardial infarction, stroke, and type 2 diabetes [2-5]. These are the leading causes of disability worldwide [6]. The consequences of metabolic disorders are often treated by healthy weight, diet and physical activities [7, 8]. So, evaluation of metabolic risk factors and the identification of population groups at risk of chronic diseases are essential for developing prevention strategies. Hence, the dynamic modeling of biological systems to describe various human diseases is of interest in recent years.

The complex network of proteins (gene products) and their biological processes mediating interactions among them in these diseases are of importance to understand. The application of protein interaction networks to available disease datasets in the public domain allows the identification of genes and their corresponding proteins. This helps the creation of sub-networks to study network properties for the classification of disease-associated genes in networks. It is found that several strategies have been employed to analyze gene networks using data for protein interactions in these conditions. However, this is a complex and a challenging task to pursue [9]. The information related to the disease mechanism gleaned using data for gene networks at a system level is critical yet it is highly convoluted. This is possible by collecting relevant data followed by cleaning such data by removing redundant information for useful yet specific knowledge establishments. This is helpful for improved data analysis followed by data integration to create a reliable model of the disease under study. Thus, gene network methods have been used to gain insights into disease mechanisms [10, 11], co-morbidity (anomalous conditions) [12, 13], protein target identification [14-16] and biomarker detection [17, 18]. The
gene network based study includes elucidation of a complex system by fragmenting them into finite components (nodes or vertices) and interactions (edges). This conceptual illustration helps in the understanding of complex molecular disturbances in diseased conditions. Therefore, it is of interest to use graph theory based pathway diagrams using pertinent co-localization information with shared domain data between MD and ARD by mediating protein-protein interaction networks to identify the genes in a common pathway among disease types, states and conditions.

Methodology:

**Disease associated gene data collection from known literature**
We gathered disease associated proteins (gene products) and or their corresponding genes related data from publically (WWW – World Wide Web) available databases such as PubMed (http://www.ncbi.nlm.nih.gov/pubmed), PubMed Central (PMC - (http://www.ncbi.nlm.nih.gov/pmc) and other open access journals maintained by several publishers across Nations. This is done through disease specific manual keyword (metabolic disorder (MD), age related disorder (ARD), relevant genes) searching, article gathering, visual scanning, reading, studying, understating, cleaning, grouping, labeling, refining, storing in simple RDBMS, and subsequent data retrieval for value addition, information enrichment and knowledge creation on the subject of the study. It should be noted that PubMed and PMC are maintained at National Centre for Biotechnology (NCBI), National Institute of Health (NIH), USA.

**Methodology**

**Disease specific network creation**
We used GeneMANIA (http://www.genemania.org/) [19] to collect data related to metabolic disorder (MD) for relevant information gathering and knowledge establishment with available graphical network diagrams retrieved from the server in this study. GeneMANIA provides data for protein-protein interactions, protein-DNA interactions and or protein-gene interactions, corresponding pathways, associated reactions, available phenotypic profiles and genes expression data with corresponding known yet characterized proteins in the network. The data thus described is reasonably representative if not comprehensive.

**Common pathway identification**
We used the pathway database (http://www.pathwaycommons.org) [20] to identify the common pathways amongst genes of metabolic disorder (MD). Pathway commons is a collection of pathway information from multiple organisms. A comprehensive collection of biological pathways from multiple sources is provided by pathway commons and represented in a common language for genes based metabolic pathway analysis.

**Pathway analysis**
We used Reactome (http://www.reactome.org/), a curated, peer reviewed and an open source pathway database for exploring the pathway knowledge for common genes [21].

**Gene localization and domain analysis**
Genes are often expressed in the same tissue and corresponding proteins can be found in the same location. Two or more genes are linked if they are expressed in the same tissue and their corresponding proteins are found in the same cellular location. Similarly, two proteins (gene products) are linked if they have the same defined (sequence and or structure) protein domain. This is completed using InterPro, SMART and Pfam facilities in the public domain. Gene localization and domain analysis are completed using the tools at GeneMANIA (http://www.genemania.org/).

**Results & Discussion:**
The hypothesis describing the mechanism leading the molecular disturbances in metabolic disorders (MD) and age-related disorders (ARD) is usually non-trivial nature. Therefore, it is important to relate a disease condition with its know yet reasonably representative associated genes for the construction of its corresponding protein-protein interaction networks. It is also of critical importance to identify genes and or their protein products that share common pathways between different disease conditions (e.g. MD and ARD). Therefore, we used text-mining (keyword searching) techniques to identify genes associated with MD as shown in Figure 1. The text-mining analysis searched for respective known genes associated with 36 metabolic diseases (Table 1). Thus, genes known to be associated with each described metabolic disorder is listed.

The listed data were further analyzed to identify key genes associated with various diseases for network analysis and evaluation (Figure 2). This exercise identified five common pathways with predicted genes for possible disease association from the network analysis (Figure 3). Subsequently, common pathway genes with specific diseases are thus summarized (Table 2). This data shows that APOB, LDLR, APOE, LIPC genes are found in common pathways for lipid digestion, mobilization, and transport. Results show that two genes APOE, LIPC were not in MD but they shared the common pathway with APOB, LDLR responsible for Glycogen storage Disease type 0 (GSD). These two predicted genes are known to cause age-related macular degeneration (AMD) [22, 23]. The BCKDHA, DBT, DLD and BCKDHB genes were found in the pathways of branched-chain amino acid catabolism and...
metabolism of amino acids and its derivatives. However, the predicted DLD gene was known to be associated with skin photo aging [24]. The other three genes (BCKDHA, DBT, and BCKDHB) are the causative agent of MD described as Male Syrup Urine Disease (MSUD) as given in Table 2.

Table 1: Description of common genes in pathways associated with specific diseases.

| Metabolic disorders | Responsible genes |
|---------------------|-------------------|
| Diabetes Type 1     | HLA DQA1, DQB1, DRB1 |
| Phenyl-keto-nuria   | TCF7L2, PPARG, KCNJ11, NOTCH2, WFS1, IGF2BP2, SLC30A8, JAZF1, HHEX |
| Glucose galactose mal-absorption | SGLT1 |
| Tyrosine-mia        | PAH |
| Alkaptonuria        | CBS, MTHFR, MTR, MTRR, MMADHC |
| Maple syrup urine disease | BCKDHA, BCKDHB, DBT |
| Propionic acid-emia | PCCA, PCCB |
| Methyl-malonic acid-emia | MUT, MAA, MMAB, MMADHC, MCEE |
| Hyper-cholesterol-emia | APOB, LDLR, LDLRAP1, PCSK9 |
| Glucose galactose mal-absorption Type 0 | GYS1, GYS2 |
| Type 1              | G6PC, SLC37A4 |
| Type 2              | GAA (Pomp disease) |
| Type 3              | AGL |
| Type 4              | GBE1 |
| Type 5              | PYGM |
| Type 6              | PYGL |
| Type 7              | PEKMT |
| Type 9              | PHKA1, PHKA2, PHKB, PHKG2 |
| Type 10             | PGAM2 (Phospho-glycerate mutase deficiency) |
| Galactose-mia Type 0 | GALE, GALK1, GALT |
| Type Ic             | ALG6 |
| Type iii            | CCDC5 |
| Lesch-Nyhan syndrome | HPRF1 |
| Gaucher disease     | GBA |
| Tay-Sachs disease   | HEXA |
| Fabry disease       | GLA |
| Hunter syndrome     | IDS |
| Sanfilippo syndrome | GN5, HGSNAT, NAGLU, SGSH |
| Moroteaux-Lamy syndrome | ARSB |
| Morquio syndrome    | GALNS, GLB1 |
| Refsum disease      | PEX7, PHYH |
| Hemo-chromatosis    | HAMP, HFE, HFE2, SLC40A1, TFR28 |

Figure 2: A network of genes associated with the metabolic disorder (MD) is illustrated in this diagram. It should be noted that genes that are already known to be associated with the disorder is shown in the figure. Black nodes indicate known MD genes and gray nodes indicate genes that are predicted to be associated with the disorder.
Table 2: Description of manually gleaned common pathway genes with known specific diseases

| Genes | Pathway ID R-HAS (prefix) | Pathways | Known disease genes | Common pathway (predicted) related | Reference |
|-------|---------------------------|----------|---------------------|------------------------------------|-----------|
| APOB; LDLR; APOE; LIPC | 73923 | Lipid digestion, mobilization and transport | GSD | APOB | APOE | AMD | [22] |
| BCKDHA; DBT; DLD; BCKDHB | 70895 | Branched chain amino acid catabolism | MSUD | BCKDHA; BCKDHB; DBT | DLD | SP | [24] |
| PYGM; PHKB; PHKG1; PHKA1 | 71291 | Metabolism of amino acids and its derivatives | | | | | |
| HEPH; CP; SLC40A1 | 1430728 | Metal ion SLC transporters | HC | SLC40A1 | HEPH | AMD | [26] |
| SLC40A1 | 425410 | Transport of sugars (glucose), bile salts, organic acids, metal ions and amine compounds | | | | | |
| 425366 | Trans-membrane transport of small molecules | | | | | | |
| ABCC8; KCNJ11 | 1296025 | ATP sensitive K+ channels | Type 2 diabetes | KCNJ11 | ABCC8 | Ageing | [27] |
| 1296065 | Inwardly rectifying K+ channels | | | | | | |
| 382556 | ABC proteins mediated transport | | | | | | |
| 422356 | Insulin secretion regulation | | | | | | |
| 1296071 | K+ channels | | | | | | |
| 163685 | Energy metabolism | | | | | | |
| 112316 | Neuronal system | | | | | | |
| 382551 | Trans-membrane transport of small molecules | | | | | | |
| 1430728 | Metabolism related | | | | | | |

ID = identifier; ATP = adenosine tri-phosphate; K+ = potassium; SLC = solute carrier; ABC = ATP-binding cassette; AMD = age-related macular degeneration; HC = Hemochromatosis; GSDV = Glycogen storage disease type 5; GSDIX = Glycogen storage disease type 9; MSUD = Maple Syrup Urine Disease; GSD = Glycogen storage disease type 0; SP = skin photo aging; glyco-genolysis = glycogen breakdown

Figure 3: An illustration showing common pathways between MD and ARD is given. (A) Common pathways of MD with predicted disorder associated genes are shown. (B) Five common pathways between MD and predicted disorder genes are illustrated.
Figure 4: A network between co-localized and domain-sharing genes is shown. Black nodes indicate the common pathway of genes between MD and ARD.

The four genes (PYGM, PHKB, PHKG1, PHKA1) were found in a common pathway from the network analysis sharing glycolgenolysis (glycogen breakdown) and carbohydrate metabolism. The predicted gene PHKG1 is known to be associated with adenoid cystic carcinoma [25]. The other three genes PYGM, PHKB and PHKA1 involved in glycogen storage disease type 5 (GSDV) and glycogen storage disease type 9 (GSDIX) is also observed in the network analysis. The other set of genes (HEPH, CP and SLC40A1) were also found in common pathways of metal ion SLC transporters, iron uptake and transport, glucose transport, transport of bile salts and organic acids, metal ions and amine compounds, SLC-mediated transmembrane transport and trans-membrane transport of small molecules. It should be noted that HEPH and CP are not associated with MD but they are involved with age-related macular degeneration (AMD) [25].

Two genes (ABCC8 and KCNJ11) associated with Type 2 diabetes (Table 2) and aging share a common pathway and hence have a linkage [26]. ABCC8 and KCNJ11 shared the common pathways for ATP sensitive K⁺ channels, inwardly rectifying K⁺ channels, ABC-family proteins mediated transport, regulation of insulin secretion, integration of energy metabolism, neuronal system association, trans-membrane transport of small molecules and other metabolism as given in Table 2. Common genes associated with both MD and ARD are further processed for molecular interactions using network analysis as described in the methodology section. It is further showed that these genes shared the same domain in pathway regulation (Figure 4). They are also co-localized with each other in tissues when expressed.

It is of important to understand the specific molecular pathways unique to a disease to elucidate the difference in these pathways. Therefore, it is essential to construct a ‘linkage network’ between diseases that are inter-linked by one or more genes found associated with the diseases using simplified network diagrams. We illustrated a linkage network based on pathway data, domain information and co-localization analyses between MD and ARD in this report. Thus, a model hypothesis is proposed using interaction networks that are linked between MD and ARD (Figure 5).

Figure 5: Venn diagram showing MD (dark green) and ARD (dark blue) disorders, and common genes of MD (light green) and ARD (light blue) is shown. Set of common pathway of genes for MD (light green) and intersection (Ash) for set of common genes between MD (light green) and ARD (light blue), \((\text{MD}) \cap (\text{ARD}) = \{\text{APOE, LIPC, HEPH, CP, DLD, ABCC8}\} \) are shown.

Conclusion:
We report data for common pathway of genes responsible for metabolic disorders (MD) and age related disorders (ARD). Data shows the linkage of genes in these diseases by analyzing their co-localization and shared domains. Pathway analysis with gene regulatory network evaluation using gene circuits and module design for further analysis of gene product interactions is essential for understanding the mechanism of the disease in relation to molecular cellular biology events.
References:
[1] http://www.idf.org/metabolic-syndrome
[2] Mathieu P et al. Vasc Health Risk Manag. 2006 2: 285 [PMID: 17326334]
[3] Garg PK et al. Hypertension. 2014 63: 413 [PMID: 24191289]
[4] Olijhoek JK et al. Eur Heart J. 2004 25: 342 [PMID: 14984924]
[5] Wilson PW et al. Circulation. 2005 112: 3066 [PMID: 16275870]
[6] http://www.who.int/mediacentre/factsheets/fs310/en/
[7] Dutheil F et al. Nutr J. 2012 11: 72 [PMID: 22985437]
[8] Andersson SA et al. Scand J Med Sci Sports. 2007 17: 687 [PMID: 17331082]
[9] Sevimoglu T & Arga KY, Comput Struct Biotechnol J. 2014 11: 22 [PMID: 23579140]
[10] Goh KI et al. Proc Natl Acad Sci U S A. 2007 104: 8685 [PMID: 17502601]
[11] Yang X et al. Nat Genet. 2009 41: 415 [PMID: 19270708]
[12] Lee DS et al. Proc Natl Acad Sci U S A. 2008 105: 9880 [PMID: 18599447]
[13] Park J et al. Mol Syst Biol. 2009 5: 262 [PMID: 19357641]
[14] Yildirim MA et al. Nat Biotechnol. 2007 25: 1119 [PMID: 17921997]
[15] Winter C et al. PLoS Comput Biol. 2012 8: 5 [PMID: 22615549]
[16] Gottlieb A et al. Mol Syst Biol. 2011 7: 496 [PMID: 21654673]
[17] Taylor IW et al. Nat Biotechnol. 2009 27: 199 [PMID: 19182785]
[18] Re M & Valentini G, IEEE/ACM Trans Comput Biol Bioinf. 2013 10: 1359 [PMID: 24407295]
[19] Warde-Farley D et al. Nucleic Acids Res. 2010 38: 214 [PMID: 20576703]
[20] Cerami EG et al. Nucleic Acids Res. 2011 39: 685 [PMID: 21071392]
[21] Milacic M et al. Cancers (Basel). 2012 4: 1180 [PMID: 24213504]
[22] Klaver CC et al. Am J Hum Genet. 1998 63: 200 [PMID: 9634502]
[23] Neale BM et al. Proc Natl Acad Sci U S A. 2010 107: 7395 [PMID: 20388286]
[24] Moon E et al. J Proteomics. 2015 117: 70 [PMID: 25576851]
[25] Bell A et al. Cancer. 2011 117: 2898 [PMID: 21692051]
[26] Ramkumar HL et al. Prog Retin Eye Res. 2010 29: 169 [PMID: 20206286]
[27] Efferth T, Ageing Res Rev. 2003 2: 11 [PMID: 12437993]

License statement: This is an Open Access article which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly credited. This is distributed under the terms of the Creative Commons Attribution License.