The MicroRNA Interaction Network of Lipid Diseases

Abdul H. Kandhro 1,2, Watshara Shoombuatong 2, Chanin Nantasenamat 2, Virapong Prachayasittikul 3 and Pornlada Nuchnoi 1, 4

1 Center for Research and Innovation, Faculty of Medical Technology, Mahidol University, Bangkok, Thailand, 2 Center of Data Mining and Biomedical Informatics, Faculty of Medical Technology, Mahidol University, Bangkok, Thailand, 3 Department of Clinical Microbiology and Applied Technology, Faculty of Medical Technology, Mahidol University, Bangkok, Thailand, 4 Department of Clinical Microscopy, Faculty of Medical Technology, Mahidol University, Bangkok, Thailand

Background: Dyslipidemia is one of the major forms of lipid disorder, characterized by increased triglycerides (TGs), increased low-density lipoprotein-cholesterol (LDL-C), and decreased high-density lipoprotein-cholesterol (HDL-C) levels in blood. Recently, MicroRNAs (miRNAs) have been reported to involve in various biological processes; their potential usage being a biomarkers and in diagnosis of various diseases. Computational approaches including text mining have been used recently to analyze abstracts from the public databases to observe the relationships/associations between the biological molecules, miRNAs, and disease phenotypes.

Materials and Methods: In the present study, significance of text mined extracted pair associations (miRNA-lipid disease) were estimated by one-sided Fisher’s exact test. The top 20 significant miRNA-disease associations were visualized on Cytoscape. The CyTargetLinker plug-in tool on Cytoscape was used to extend the network and predicts new miRNA target genes. The Biological Networks Gene Ontology (BiNGO) plug-in tool on Cytoscape was used to retrieve gene ontology (GO) annotations for the targeted genes.

Results: We retrieved 227 miRNA-lipid disease associations including 148 miRNAs. The top 20 significant miRNAs analysis on CyTargetLinker provides defined, predicted and validated gene targets, further targeted genes analyzed by BiNGO showed targeted genes were significantly associated with lipid, cholesterol, apolipoprotein, and fatty acids GO terms.

Conclusion: We are the first to provide a reliable miRNA-lipid disease association network based on text mining. This could help future experimental studies that aim to validate predicted gene targets.

Keywords: microRNA, text mining, interaction network, lipid diseases, dyslipidemia

INTRODUCTION

Dyslipidemia is a common form of lipid disorder; characterized by increased levels of TGs, increased LDL-C, and decreased level of HDL-C. The low levels of HDL-C and high levels of LDL-C are the most imperative factors for the development of cardiovascular disease (CVD), especially ischemic heart disease and stroke (Meagher, 2004). The liver is the major organ where cholesterol,
lipid, and lipoprotein synthesis and metabolism taking place (Min et al., 2012). Recently, miRNAs have been reported to modulate these processes (Esau et al., 2006; Moore et al., 2010; Vickers et al., 2013). The miRNA-122 is identified for the involvement of the regulation of lipid metabolism (Esau et al., 2006). A recent study that shows miRNA-27b targeted to 27 of 151 lipid-associated genes. This therefore indicates that miRNA-27b serves as a key molecule for post-transcriptional hub of lipid metabolism genes. In mice model, GPAM is one of the key lipid metabolism gene targeted by miR-27b. The up-regulation of hepatic miR-27b is associated with decrease GPAM mRNA and plasma triglyceride (Vickers et al., 2013).

Moreover, miRNA-33a and miRNA-33b have been extensively identified to be involved in cholesterol and lipid homeostasis. miRNA33a and miRNA-33b are experimentally characterized to be located in intronic regions of sterol regulatory element binding protein-2 and 1 (SREBP2, SREBP1), respectively (Marquart et al., 2010; Najafi-Shoushtari et al., 2010; Rayner et al., 2010). The finding of hepatic miR-27b is promising for modulating the endogenous miR-27b level as effective therapeutic approach in lipid related disorder in further study.

miRNAs are evolutionary conserved and found commonly in humans, flies, plants, and viruses (Lagos-Quintana et al., 2003). Signaling proteins, metabolic enzymes, transcription factors are regulated by miRNAs. The expression levels of miRNAs have demonstrated their potential usage as biomarkers for various diseases (Lagos-Quintana et al., 2003). Although there have been advancements in miRNA profiling, the experimental process for searching disease related miRNAs is considerably expensive and time-consuming (Jiang et al., 2010, 2013). Computational approaches have been proposed to overcome these limitations and drawbacks in miRNA research. A large number of miRNA prediction softwares have been developed to predict miRNA by targeting 3’UTR of mRNA such as, miRanda (Enright et al., 2004), TargetScan (Lewis et al., 2003), and PicTar (Krek et al., 2005). The computational prediction algorithms mostly analyze the binding of miRNA at 3’UTR region of human mRNA. With the public availability of human genome information and miRNA bioinformatic tools, a large number of research publications related to miRNA in human diseases have been published, and are now in databases such as, PubMed and Scopus. Text-mining is one of the promising tools for depicting the body of knowledge from the literature. Naeem et al. (2010) demonstrated the use of co-occurring based text mining method for elucidating miRNA-gene association. Murray et al. (2010) uncovered human miRNA-target interactome (microRNAome), using natural language processing (NLP) based text-mining, network analysis, and ontological enrichment methods. Goh et al. (2007) described the involvement of 176 miRNAs and their target genes in the controlling of 368 OMIM disorders using human disease network. There is a need for data reduction methods i.e., text-mining, that utilize validated miRNA-disease associations from experimental published abstracts which indicates most significantly disease related miRNAs for experimental study.

We constructed a miRNA-lipid disease association network using computational approaches; including text-mining approach with miRNA bioinformatic tools. This is the first study that delineates the interacting network of miRNAs and the target genes in human lipid disorders.

**MATERIALS AND METHODS**

**Data Collection**

In this study, data was collected from January 1, 2000 to December 31, 2013. A total of 730 abstracts were collected from publicly available databases like PubMed and Scopus by using keywords for lipid diseases/identifier and miRNA terms. Figure 1 shows the workflow diagram of the present study.

In the present study, the text mining framework was divided into five main steps, namely:

(i) Information Retrieval (IR)

(ii) Information Extraction (IE) & Scoring

(iii) MicroRNA-disease Network Construction

(iv) Extension of network to predict new miRNA targets

(v) Identification of Gene Ontology (GO) terms for predicted targets.

Figure 2 shows the text-mining framework into five steps

**Information Retrieval (IR)**

The fundamental assumption in the field of text mining is that co-occurrence means association. Based on the co-occurrence assumption, the associations between different miRNAs and lipid diseases or identifiers were determined. If a particular miRNA and lipid disease or identifier were mentioned in the same abstract, we assumed that they co-occurred and were associated.

Information retrieval highly relies on the keyword recognition, which is the miRNA name and disease or disease identifier groups, then the set of keywords used to search within the databases and retrieve the keywords containing abstracts. We used different keyword terms for miRNA prefixes, because of the different patterns of recognition for their names such as: “MicroRNA,” “MiRNA,” “mir,” prefixed species as “hsa-miR-1,” as precursor “pre-miR-1,” as loci or variant “miR-1a-1.” Other variants like “lin-4” and “let-7,” as an abbreviation more than one miRNA “miR-221/222” and “miR-15 – 16,” However, for disease or identifier groups, we selected the four common lipid diseases and seven identifiers which are related to the 4 common diseases. The common diseases are Dyslipidemia, Hyperlipidemia, Hypercholesterolemia, and Hypertriglyceridemia, while the 7 identifiers such as, HDL-Cholesterol, LDL, LDL-Cholesterol, LDL, Triglyceride, Low HDL-C, and High LDL-C, Low HDL-C. We made a pair of miRNA+disease or miRNA+identifiers for searching abstracts from both PubMed and Scopus databases from January 1, 2000 to December 31, 2013, and saved these as a notepad file (for example: “MicroRNA and Dyslipidemia,” “MiRNA and Dyslipidemia,” “miR and dyslipidemia,” “hsa-miR and Dyslipidemia,” and similarly other miRNA recognition terms).

**Information Extraction (IE) and Scoring**

The simplest form of the approach used in our study depend on the relation of keywords between abstracts is an association based on the co-occurrence of the keywords in the text. When two
keywords are frequently mentioned in the abstract, an association relation between keywords is inferred. By using EditPlus software (https://www.editplus.com/), which is used for manual text mining or information extraction (IE) from reference abstracts.

The significance level of extracted miRNA-disease association pairs were computed by one-sided Fisher’s exact tests (Fisher, 1922). The $P$-value of Fisher’s exact tests (Fisher, 1922) was calculated based on hypergeometric distribution, as follows:

$$P = \frac{(a+b)! (c+d)! (a+c)! (c+d)!}{a!b!c!d!n!}$$

where $n$ is denoted the total number of abstracts included in text mining;

- $a$ is the True positive (TP) which represents the number of abstracts that contain both the miRNA and disease;
- $b$ is the False positive (FP) which represents the number of abstracts that contain only miRNA;
- $c$ is the False negative (FN) which represents the number of abstracts that contain only the disease/identifier;
- $d$ is the True negative (TN) which represents the number of abstracts that don’t contain either terms.

The $P$-value, which determines whether a miRNA and disease have a link, is considered significant as $\leq 0.05$.

(iii) MicroRNA-Disease Network Construction

One of the commonly used framework to visualize and analyze biological network is Cytoscape (Shannon et al., 2003). It provides functionality for representation and integration of biomolecular network models. In present study, we constructed the bipartite network by mapping pairs of miRNA-disease associations based on $P$-values, and visualized the network by Cytoscape v3.2.0. Here, the disease groups attributed to a node,
and miRNAs attributes to an edge. The interaction between disease and miRNA weighted with corresponding $P$-values. Each edge in the network connects a miRNA and one of its corresponding one or more than one disease group, similarly each disease group corresponds one miRNA or more than one miRNAs. Thus, resulting constructed miRNA-disease association network provides information on whether miRNA is associated with a disease.

(iv) Extension of Network (New miRNA Targets Predictions)

Cytoscape has a modular structure and extension of networks with additional functionalities is possible through apps (formerly known as plugins). Currently, few Cytoscape apps are available that either extend networks with other types of molecular interaction data or focus on one specific type of regulatory interaction. A new Cytoscape app, CyTargetLinker (Kutmon et al., 2013) allows users to build regulatory interaction networks, and allow their inclusion in the network analysis process.

We used CyTargetLinker v3.0.1 to validate and predict miRNA target interactions (MTIs) and visualize them in a graphical way by extension of the network. A regulatory interaction network (RegIN) is a network containing regulatory interactions often derived from online interaction databases. To construct a RegIN with CyTargetLinker on Cytoscape, we obtained Homo sapiens MTIs from one experimentally validated database miRTarBase v4.4, which includes 20,942 MTIs, and from two predicted miRNA databases; MicroCosm v5.0, which includes 541,039 MTIs and TargetScan v6.2, which includes 511,040 MTIs. The networks are stored in XGMML (the eXtensible Graph Markup and Modeling Language) format, which is supported by Cytoscape. Each regulatory interaction consists of two nodes, a source (regulatory component) and target biomolecule, connected through one directed edge. The CyTargetLinker website http://projects.bigcat.unimaas.nl/cytargetlinker/regins provides a collection of RegINs for different species and interaction types. Figure 3 shows the workflow diagram of CyTargetLinker.

In present study, top 20 significant miRNAs selected for extension of network by CyTargetLinker. The top 20 significant
miRNAs then divided into four sets (each set contains five miRNAs) with miRBase accession numbers and used as input file for CyTargetLinker.

- The first step is to load input file on Cytoscape.
- In the second step, CyTargetLinker integrates and extends the network process is started. In the dialogue box before network extension user can add either targets or regulators or both as default. As a result CyTargetLinker extracts the RINs from the provided MTIs. After the extension of network, CyTargetLinker fixes different colors on each edge for targets, regulators, and MTIs. We can see the detail on control panel, where color selection and number of MTIs for each databases is listed.
- In the fourth step the network MTIs can be visualized by adopting the hide/show and/or overlap threshold function. The function of hide/show key is to enable the temporary removal of specific MTIs and showing only the interactions from a subset of loaded MTIs. However, the function of overlap threshold key is to show only the interactions that are supported by a defined number of MTIs or more. After the most targeted MTIs visualized in Cytoscape/CyTargetLinker RIN network, the targeted MTIs were used to retrieve the GO for identifying their biological processes. For GO, another Cytoscape Plug in tool, BiNGO (Maere et al., 2005) was used.

(v) Gene Ontology
The BiNGO v3.0.3 (Maere et al., 2005) is a Cytoscape plugin tool used to retrieve the GO annotations for the targeted genes identified with CyTargetLinker. Figure 4 shows the workflow diagram of BiNGO Gene ontology analysis.

By using the input list of targeted genes; BiNGO accesses the overrepresentation of GO categories in a subgraph of a biological network, which is visualized on Cytoscape. The enrichment of GO terms in the targeted genes was evaluated with a right-sided hypergeometric statistical analysis. The hypergeometric test $P$-value was set to $\leq 0.05$, and the Benjamini and Hochberg correction was applied to provide strong control over the false discovery rate under positive regression dependency of the test statistics. After statistical analysis, the GO hierarchy was visualized as overrepresented GO categories.

The main advantages of BiNGO are:

(i) It supports GOSlim ontologies (Consortium, 2004),
(ii) It offers enormous flexibility in the use of ontologies and annotations,
(iii) It can be integrated with a range of molecular networks including protein-protein interactions or transcriptional co-regulation networks,
(iv) It allows networks to be modified, viewed and analyzed in various ways on Cytoscape.

RESULTS

MicroRNA-Lipid Disease Association Analysis
In the present study, the associations were identified by co-occurrence-based manual text-mining approach and significance was measured by one-sided Fisher’s exact $P$-values. Significant associations were used to construct the network on Cytoscape. By processing 730 publications, we recorded 227 pairs of miRNA-lipid disease associations. Among these associations, there are 148 miRNAs and 09 (04 diseases, 05 identifiers) groups involved. Table 1 gives an overview of the number of miRNAs, diseases, miRNA- lipid disease associations, and a number of papers.

MicroRNA-Lipid Disease Association Network Construction and Visualization
The construction of bipartite network of miRNA-disease associations based on the $P$-values. The $P$-values of each association was computed by one-sided Fisher's exact test, and were calculated based on hypergeometric distribution. The bipartite network consists of 157 nodes (corresponding to disease/identifier and miRNAs) and 227 edges (corresponding to miRNA-disease associations). We prioritized 148 miRNAs in 4 diseases and 5 identifier groups and all miRNA-disease associations shown in Figure 5. The top 20 significant association network constructed which is based on edge-weighted $P$-values, shown in Figure 6. The higher significant $P$-values correspond to more thicker edges between each pair. The higher strength is shown in HDL-Cholesterol and Triglyceride group paired miRNAs.

The top 20 significant miRNAs-disease pairs selected with a number of papers by applying the one-sided Fisher's exact $P$-values. Table 2 shows the top 20 significant associations with $P$-values. From the nine disease/identifier groups, only four groups show the higher number of pairs including, HDL-Cholesterol, Triglyceride, Hyperlipidemia, LDL group as seen in Table 2. The miRNA-33 family have shows a higher number of papers and mostly paired in the HDL-Cholesterol group.

Construction of Regulatory Interaction Network (New miRNA Target Predictions) by CyTargetLinker on Cytoscape
The CyTargetLinker on Cytoscape used to extend the RIN network, which augments user knowledge about the new miRNA target predictions that could be used for further experimental studies. To get better insight from present top 20 miRNA target predictions and extend our RegIN, we used CyTargetLinker Plug in application on Cytoscape. We obtained experimentally validated MTIs by miRTarBase database and predicted MTIs by MicroCosm and TargetScan databases, which are described in Table 3, where the number of nodes, edges, number of validated, and predicted targets is listed as follows:

We found that CyTargetLinker provides quick and extensive enrichment of biological network with regulatory information. At threshold 3 functionality on the control panel, we observed the defined number of targeted genes that provide regulatory
CyTargetLinker (RINs network) workflow diagram

miRNA input in Cytoscape

hsa-miR-122  hsa-miR-33  hsa-miR-144  hsa-miR-96  hsa-miR-145

Download RINs

Regulatory Interaction (RINs)
1. miRTarBase v4.4 (20,942 MTIs)
2. MicroCosm v5.0, 541,039 MTIs
3. TargetScan v6.2, 511,040 MTIs

Network visualization (node selection, RINs and direction)

CyTargetLinker Control panel

STEP-1

STEP-2

STEP-3

STEP-4

At hide threshold

Validated MTIs

At overlap threshold

Overlapped MTIs

FIGURE 3 | Workflow diagram of CyTargetLinker regulatory interaction network.
interactions of validated and predicted MTIs, and shown in Table 4 and Figure 7.

The number of defined targeted genes by miRNAs is shown in Figure 7. The higher to lower number of MTIs are; miRNA-155 targeted 31 genes, miRNA-103 targeted 11 genes, and miRNA-200c targeted 10 genes. The targeted genes are derived from validated and predicted MTIs databases. The defined numbers of each miRNA target genes are listed in Table 5.

**Gene Ontology Analysis**

Besides the RegIN information, next step to obtain biological functions of targeted genes for further understanding the
biological role of the gene. To gain further insight into the molecular aspects of above listed miRNAs signature in lipid disorders, we investigated the GO for biological, cellular, and molecular processes associated with a set of predicted and validated targeted genes by miRNAs. Surprisingly, we found more than 90 related GO terms shown in Supplementary Material among which more than 20 GO terms were significantly associated with lipids, cholesterol, fatty acid, apolipoproteins, sterol, and insulin and shown in Table 6. When we narrowed down searching on molecular and cellular processes related GO terms, at this stage the GO terms were associated with gene activity, negative and positive regulation of metabolic process, regulation of biological process, metabolic processes, cellular metabolic processes, transportation of lipids, storage of lipids, cholesterol efflux, and macromolecular biosynthetic processes.

Taken together, integrated results of CyTargetLinker target genes analyzed by BiNGO GO terms, suggested that the targeted genes are associated with lipid, cholesterol, lipoprotein, fatty acid, and insulin that significantly involved in their biological, metabolic, and cellular processes. It could be elucidated for their respective pathogenic role and molecular mechanism of action in lipid, cholesterol, lipoprotein, fatty acid, and insulin disorders. In addition, the BiNGO results also highlighted GO terms in cell cycle, cell differentiation, apoptosis/cell death, signaling pathways, protein and carbohydrate metabolisms, immune system, and neuronal metabolism. Therefore, annotated GO terms could help in examining the relationships between the miRNAs and their targets in cancers, metabolic diseases of carbohydrate and proteins, immune diseases, and neurological diseases.

**Performance Evaluation**

We further compared our study with three other existing databases. For this purpose, we manually checked, confirmed and compared Top 20 miRNAs (of present study) with existing databases such as miR2Disease, miRiaD, and HMDD. We have found that most of the associations missed in miR2Disease and HMDD databases, while miRiaD database missed only few associations shown in Supplementary Table 2. However, Supplementary Figure 1 shows the comparison of our study with miRiaD database. The failure of association in other databases might be due to these databases present most of the miRNAs associations with cancers, while few miRNAs associated with metabolic and other diseases.

**DISCUSSION**

The present study is applicable to signify associations between miRNAs and common lipid diseases, where the significant associations were used to visualize and construct the RegIN. The text-mining approach is helpful for extracting the information from huge literature to small subset of extracted information, which is then used for potential knowledge discovery. Hence, we may call the subset information as “literature verified” information.

We are first to provide independently the miRNA-lipid disease associations with network visualization, extension of network for predicted/validated target genes with their associated GO terms. Present study possesses limited number of publication abstracts, although we retrieved abstracts by January 1, 2000 to December 31, 2013. By processing 730 abstracts, we found 227 pairs of miRNA-lipid disease associations, and prioritized 148 miRNAs in nine disease/identifier groups. The major reasons for limited numbers of publications are (a) failure of experimental studies for the discovery and identification of miRNA genes and their targets (Grosswendt et al., 2014), (b) expensive experimental methods for identifying disease related miRNAs and shown low sensitivity & specificity (Jiang et al., 2010, 2013). The co-occurrence based text-mining approach adopted in other studies like, Naeem et al. (2010) for identifying mRNA and genes co-occurring in abstracts; Lu et al. (2008) for identifying miRNA-disease associations; and Jiang et al. (2009) for identifying mRNA-disease relationships. Although, their approaches were effective on limited or low number of publications (100 by Lu et al., 2008 and 600 by Jiang et al., 2009), but not for high scale of text-mining as the number of miRNA research publications increases regularly. In addition, high false positive rate found in their studies, which may lead to poor resolution of miRNA-targets.

We found higher strength of miRNAs in HDL-Cholesterol and Triglyceride groups with higher number of abstracts co-occurring miRNA-33 family. The expression of intronic miRNA-33 family (miR-33a and miR-33b) are from the sterol regulatory element-binding protein (SREBP) transcription factors, which are known to be involved in cholesterol/lipid homeostasis, and many cholesterologenic/lipogenic genes like LDL-Receptor, 3-hydroxy-3-methylglutaryl coenzyme A reductase (HMGC), fatty acid synthase (FAS) (Brown and Goldstein, 1997, 2009; Horton et al., 2002; Osborne and Espenshade, 2009). The conserved target for miR-33a and miR-33b is adenosine triphosphate binding cassette A1 (ABCA1) cholesterol transporter. ABCA1 helps to transport intracellular cholesterol from liver to apolipoprotein A-1 (apo-A-1) for the synthesis of HDL-C (Maxfield and Tabas, 2005; Wang and Rader, 2007; Tall et al., 2008). At the same extent, by reverse cholesterol transport (RCT) pathway, HDL-C transfers from peripheral tissues/macrophages back to the liver for processing and excretion into bile and feces (Rader et al., 2009). Both increased TGs and decreased HDL-C levels are the characteristics of dyslipidemia, found in insulin-resistant subjects, while the low HDL-C and high TGs level are the hallmarks of atherogenic dyslipidemia both in diabetic and non-diabetic populations (Group, 1997; Jeppesen et al., 1997; Hermans et al., 2012).
FIGURE 5 | All 227 miRNA-lipid disease associations by \( P \)-values. Red circles and green circles represent miRNAs and diseases, blue square represents lipid diseases/identifiers, respectively, according to the number of corresponding text mined annotated papers. Each linked pair represents a miRNA-disease association with edge-weighted measurement by \( P \)-values to visualize the strength of the miRNA-disease association. The miRNAs either connect one disease or more than one disease; it shows that a single miRNA or group of miRNAs may be involved with one or more than one disease.
By analyzing the top 20 significant miRNAs for the prediction of predicted and validated target genes by CyTargetLinker on Cytoscape, we found the higher number of defined targeted genes as shown in Figure 7. The higher to lower number of targeted genes by miRNAs listed in Table 5 such as, miRNA-155 targeted 31 genes, miRNA-103 targeted 11 genes, and miRNA-200c targeted 10 genes.

Further, targeted genes analyzed by BiNGO for GO annotation, we found more than 90 related GO terms listed in Supplementary Material, among which more than 20 GO terms were significantly associated with lipids, cholesterol, fatty acid, apolipoproteins, sterol, and insulin listed in Table 6. Taking together the defined number of targeted genes by CyTargetLinker with GO terms, suggested that validated and predicted target genes could be regulated in vivo by these significant miRNAs in lipid, cholesterol and fatty acid metabolism and associated metabolic diseases. In addition, the miRNAs may be regulated target genes in other non-lipid disorders specially cancers, neurodegenerative disorders, metabolic disorders; and several biological, cellular, and molecular impaired functions. Therefore, for future studies annotated GO terms could help in examining the relationships between the miRNAs and their targets in cancers, metabolic diseases of carbohydrate and proteins, immune diseases and neurological diseases.
FIGURE 7 | Defined number of predicted and validated target genes from Top 20 significant miRNAs. Among the 20 significant miRNAs, only 18 showed a defined number of targets at threshold 3 functionality modes on CyTargetLinker and Cytoscape. The miRNAs and their accession numbers are in red circles; pink hexagonal shapes indicate their regulatory interaction (targeted mRNAs/genes). The different colors of arrows show regulatory interactions for the identification based on miRNA validated and predicted databases (miRTarBase, MicroCosm, and TargetScan). Each miRNA targets mRNA/genes and the high number of such targets is shown in (C) where miRNA-155 has 31 targets and miRNA-200c has 11. In (D) miRNA-103 has 11 targets. Single targets also were found as shown in (A) for miRNA-144 and (B) for miRNA-33b and miRNA-185.
**CONCLUSION**

To the best of our knowledge, this study represents the first study to provide reliable miRNA-lipid disease association network based on text-mining method. We extracted 227 miRNA-lipid disease associations between 148 miRNAs and nine common lipid diseases/identifiers from bulk published data. In the present study significant groups such as, HDL-C, dyslipidemia and triglyceride should be evaluated further for identifying the complex involvement of miRNAs and disease development. We also constructed extended RegIN from top 20 significant text-mined miRNAs using CyTargetLinker on Cytoscape that

**TABLE 5** | Top 20 significant miRNAs and their predicted and validated targets by CyTargetLinker extension network analysis.

| miRNAs    | Targets                                    |
|-----------|--------------------------------------------|
| hsa-miR-122 | IGFR1, GPR172A, CCNG1, G6PC3, P4HA1, ST6GALNAC4 |
| hsa-miR-145 | IGFR1, NEDD9, SOCS7, C6orf115, SRCAP1, CLINT1, RTKN, C11orf9 |
| hsa-miR-33, hsa-miR-33a | YWHAH, PM3, ABCA1, CROT, BTBD2, NPC1, PIM1 |
| hsa-miR-144 | PFA |
| hsa-miR-96  | PROK2, TSPAN14, DDX3, FOXO3A |
| hsa-miR-223 | HSP90B1, NFA |
| hsa-miR-185 | CDC42 |
| hsa-miR-33b | ABCA1 |
| hsa-miR-155 | SDCBP, VPS18, CUL1, TRIM32, MYO10, SAP30L, RCN2, CEBPB, SOCS1, MBNL3, SMARC4, PSIP1, TP53INP1, IKKE, DET1, DHX40, MAP3K10, WEE1, HBP1, RAP1A, SP1, GNAS, KIAA1715, KIAA1411, FAT104A, TRPKB, HIFEP2, CDC73, CARHSP1, JARID2, PHC2 |
| hsa-miR-200c | BAP1, EP300, MC1R, ZEB1, IKBKB, NTF3, UBE2I, ZEB2, TBK1, FN1 |
| hsa-miR-122a | CCNG1, G6PC3, ST6GALNAC4, IGFR1, P4HA1, GPR172A |
| hsa-miR-29b  | DMN13A, COL3A1, TBX21, COL5A3, DUSP2, FOS, MYCN, DNAJB11 |
| hsa-miR-146a | TRAF6, IRK1 |
| hsa-miR-103  | ZYX, YTHDC1, ZBTB8, PD4, RNMT, CCNE1, VPS4A, TLK1, ANX2, NDRG5, VSP1, WNK3A |
| hsa-miR-375  | ELA4, AB2, LDB4, PHLDA1, RASD1 |
| hsa-miR-126  | IRS1, EGF7L, RGS3, PLK2 |
| hsa-miR-30c  | UBE2I, CUL2, SOCS1, SERPIN1, TWIF1 |

**LIMITATIONS**

There are certain limitations in our study as follows:

1. Limited number of publications in PubMed and Scopus databases as well as duplicate publications. The limited numbers of publications in both databases are due to the limited number of experimental work on disease-related miRNAs owing to the inherent expensive cost and time-consuming nature of the work. As the data in Scopus is limited to only work after 1995, therefore, searching literature from PubMed is preferable to Scopus.

2. High false positive rate is found during text mining. Because, most of the abstracts were mentioned the keyword microRNA/miRNA but not mentioned lipid disease/identifier names. Therefore, abstracts should contains information of both miRNA and disease name. 

**TABLE 6** | Top significant related gene ontology terms of CyTargetLinker targeted genes analyzed by BiGO on Cytoscape.

| GO ID | Description                                      | Total genes | Partner genes | p-value   |
|-------|--------------------------------------------------|-------------|---------------|-----------|
| 9891  | Positive regulation of biosynthetic process       | 55          | 12            | 1.83E-06  |
| 31324 | Negative regulation of cellular metabolic process | 55          | 12            | 3.88E-06  |
| 9892  | Negative regulation of metabolic process          | 55          | 12            | 1.02E-05  |
| 44260 | Cellular macromolecule metabolic process          | 55          | 25            | 1.33E-05  |
| 15248 | Sterol transporter activity                        | 20          | 2             | 6.56E-06  |
| 5159  | Insulin-like growth factor receptor binding       | 24          | 2             | 1.57E-04  |
| 46627 | Negative regulation of insulin receptor           | 24          | 2             | 2.34E-04  |

3344 Cholesterol efflux
30226 Apolipoprotein receptor activity
8286 Insulin receptor signaling pathway
30301 Cholesterol transport
42632 Cholesterol homeostasis
55092 Sterol homeostasis
34188 Apolipoprotein A-I receptor activity
6629 Lipid metabolic process
55088 Lipid homeostasis
5319 Lipid transporter activity
10887 Negative regulation of cholesterol storage
5899 Insulin receptor complex
32869 Cellular response to insulin stimulus
34186 Apolipoprotein A-I binding
10875 Positive regulation of cholesterol efflux
32365 Intracellular lipid transport
10874 Regulation of cholesterol efflux
32373 Positive regulation of sterol transport
32376 Positive regulation of cholesterol transport
34204 Lipid translocation
10885 Regulation of cholesterol storage
34380 High-density lipoprotein particle assembly
43559 Insulin binding
10888 Negative regulation of lipid storage
70325 Lipoprotein receptor binding
31325 Positive regulation of cellular metabolic process
32868 Response to insulin stimulus
9893 Positive regulation of metabolic process
32370 Positive regulation of lipid transport
6689 Lipid transport
34185 Apolipoprotein binding
10876 Lipid localization
16071 mRNA metabolic process
32368 Regulation of lipid transport
9892 Negative regulation of metabolic process
20 2 2.34E-04
20 2 2.49E-04
8 1 4.50E-04
20 2 7.39E-04
20 2 8.68E-04
20 2 1.10E-03
20 2 1.10E-03
20 1 1.12E-03
20 5 1.93E-03
20 2 2.18E-03
20 2 2.32E-03
8 1 2.70E-03
24 1 2.70E-03
20 2 3.17E-03
20 1 3.37E-03
8 1 3.59E-03
8 1 3.59E-03
8 1 3.99E-03
8 1 3.99E-03
8 1 4.04E-03
8 1 4.04E-03
8 1 4.04E-03
20 1 4.49E-03
8 1 5.38E-03
20 2 5.61E-03
20 2 5.61E-03
20 1 5.61E-03
8 1 7.17E-03
8 1 6.28E-03
8 1 6.28E-03
20 1 6.56E-03
20 1 6.56E-03
20 1 6.56E-03
8 1 8.31E-03
20 2 8.32E-03
8 1 8.51E-03
20 1 8.51E-03
20 1 8.51E-03
20 1 8.51E-03
20 1 8.51E-03
20 1 8.51E-03
20 1 8.51E-03
20 1 8.51E-03
20 1 8.51E-03
20 1 8.51E-03
20 1 8.51E-03
20 1 8.51E-03
20 1 8.51E-03
20 1 8.51E-03
20 1 8.51E-03
20 1 8.51E-03
20 1 8.51E-03
20 1 8.51E-03
20 1 8.51E-03
20 1 8.51E-03
20 1 8.51E-03
20 1 8.51E-03
20 1 8.51E-03
20 1 8.51E-03
20 1 8.51E-03
20 1 8.51E-03
20 1 8.51E-03
20 1 8.51E-03
20 1 8.51E-03
20 1 8.51E-03
20 1 8.51E-03
20 1 8.51E-03
20 1 8.51E-03
20 1 8.51E-03
20 1 8.51E-03
20 1 8.51E-03
20 1 8.51E-03
20 1 8.51E-03
20 1 8.51E-03
20 1 8.51E-03
20 1 8.51E-03
provides experimentally validated and predicted miRNA gene targets. Further, these miRNA gene targets are involved in the regulation of lipid, cholesterol, lipoprotein, and fatty acid biological processes, which are confirmed by BiNGO analysis on Cytoscape.

The current study sets the groundwork for future experimental studies to validate the targeted miRNAs’ genes, since they have been predicted with CyTargetLinker but not experimentally validated. Future experimental studies could walk around the biological functions and primary molecular mechanism of miRNAs in the development, progression, diagnosis and prognosis of lipid and cholesterol, lipoprotein, and fatty acid disorders.

AUTHOR CONTRIBUTIONS

AK is a Ph. D. student and made substantial contributions in analyzing and interpreting the research results as well as manuscript preparation. WS is a co-advisor and verified the analysis and guidance specifically in statistics. PN is a major advisor and participated in study design, preparation and reviewing of the drafted manuscript. CN and VP gave a critical review of the manuscript.

FUNDING

This work was funded by the Faculty of Medical Technology, Mahidol University, Thailand and constitute a partial fulfillment of the requirement of Ph.D. dissertation of AK and the Thailand Research Fund (TRF), the Office of Higher Education Commission (OHEC) to PN (MRG5480062).

ACKNOWLEDGMENTS

The authors would like to acknowledge Dr. Arthur Brown for the language editing for this work.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fgene.2017.00116/full#supplementary-material

REFERENCES

Brown, M. S., and Goldstein, J. L. (1997). The SREBP pathway: regulation of cholesterol metabolism by proteolysis of a membrane-bound transcription factor. Cell 89, 331–340. doi: 10.1016/S0092-8674(00)80213-5

Brown, M. S., and Goldstein, J. L. (2009). Cholesterol feedback: from Schoenheimer’s bottle to Scap’s MELADL. J. Lipid Res. 50, S15–S27. doi: 10.1194/jlr.R800054-JLR200

Consortium, G. O. (2004). The Gene Ontology (GO) database and informatics resource. Nucleic Acids Res. 32(Suppl. 1), D258–D261. doi: 10.1093/nar/gkh036

Enright, A. J., John, B., Gaul, U., Tuschl, T., Sander, C., and Marks, D. S. (2004). MicroRNA targets in Drosophila. Genome Biol. 5:R1. doi: 10.1186/gb-2003-5-1-r1

Esau, C., Davis, S., Murray, S. F., Yu, X. X., Pandey, S. K., Pear, M., et al. (2006). miR-122 regulation of lipid metabolism revealed by in vivo antisense targeting. Cell Metab. 3, 87–98. doi: 10.1016/j.cmet.2006.01.005

Fisher, R. A. (1922). On the mathematical foundations of theoretical statistics. Philos. Trans. R. Soc. Lond. A Mathemat. Phys. Charact. 309–368. doi: 10.1098/rsta.1922.0009

Goh, K.-I., Cusick, M. E., Valle, D., Childs, B., Vidal, M., and Barabasi, A.-L. (2007). The human disease network. Proc. Natl. Acad. Sci. U.S.A. 104, 8685–8690. doi: 10.1073/pnas.0701611104

Grosswendt, S., Filippchyk, A., Manzano, M., Klononomos, F., Schilling, M., Herzog, M., et al. (2014). Unambiguous identification of miRNA: target site interactions by different types of ligation reactions. Mol. Cell 54, 1042–1054. doi: 10.1016/j.molec.2014.03.049

Group, U. P. D. S. (1997). UK prospective diabetes study 27: plasma lipids and lipoproteins at diagnosis of NIDDM by age and sex. Diabetes Care 20, 1683–1687. doi: 10.2323/diabetescare.20.11.1683

Hermans, M. P., Ahn, S. A., and Rousseau, M. F. (2012). The atherogenic dyslipidemia ratio [log(TG)/HDL-C] is associated with residual vascular risk, beta-cell function loss and microangiopathy in type 2 diabetes females. Lipids Health Dis. 11:132. doi: 10.1186/1476-687X-11-132

Horton, J. D., Goldstein, J. L., and Brown, M. S. (2002). SREBP: activators of the complete program of cholesterol and fatty acid synthesis in the liver. J. Clin. Invest. 109, 1125–1132. doi: 10.1172/JCI021593

Jepesen, J., Hein, H. O., Suadicani, P., and Gyntelberg, F. (1997). Relation of high TG-Low HDL cholesterol and LDL cholesterol to the incidence of ischemic heart disease An 8-year follow-up in the copenhagen male study.

Jiang, Q., Gao, Y., Wang, X., Yuan, J., Li, Z., Zheng, M., et al. (2010). Optimization of disease microRNAs through a human phenome-microRNA network. BMC Syst. Biol. 4(Suppl. 1):S2. doi: 10.1186/1752-0509-4-S1-S2

Jiang, Q., Wang, G., Jin, S., Li, Y., and Wang, Y. (2013). Predicting human microRNA-disease associations based on support vector machine. Int. J. Data Min. Bioinform. 8, 282–293. doi: 10.1057/ijmb.2013.056078

Krek, A., Grün, D., Poy, M. N., Wolf, R., Rosenberg, L., Epstein, E. J., et al. (2005). Combinatorial microRNA target predictions. Nat. Genet. 37, 495–500. doi: 10.1038/ng1536

Kutmon, M., Kelder, T., Mandavijaya, P., Evelo, C. T., and Coort, S. L. (2013). CyTargetLinker: a cytoscape app to integrate regulatory interactions in network analysis. PLoS ONE 8:e82160. doi: 10.1371/journal.pone.0082160

Lagos-Quintana, M., Rauhut, R., Meyer, J., Borkhardt, A., and Tuschl, T. (2003). New microRNAs from mouse and human. RNA 9, 175–179. doi: 10.1261/rna.2146903

Lewis, B. P., Shih, I.-H., Jones-Rhoades, M. W., Bartel, D. P., and Burge, C. B. (2005). Prediction of mammalian microRNA targets. Cell 115, 787–798. doi: 10.1016/S0092-8674(03)0018-3

Lu, M., Zhang, Q., Deng, M., Miao, J., Guo, Y., Gao, W., et al. (2008). Analysis of human microRNA and disease associations. PLoS ONE 3:e3420. doi: 10.1371/journal.pone.0003420

Maere, S., Heymans, K., and Kuiper, M. (2005). BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. Bioinformatics 21, 3448–3449. doi: 10.1093/bioinformatics/bti551

Marquart, T. J., Allen, R. M., Ory, D. S., and Baldán, A. (2010). miR-33 links SREBP-2 induction to repression of sterol transporters. Proc. Natl. Acad. Sci. U.S.A. 107, 12228–12232. doi: 10.1073/pnas.1005191107

Maxfield, F. R., and Tabas, I. (2005). Role of cholesterol and lipid organization in disease. Nature 438, 612–621. doi: 10.1038/nature03439

Meagher, E. A. (2004). Addressing cardiovascular disease in women: focus on dyslipidemia. J. Am. Board Fam. Pract. 17, 424–437. doi: 10.3122/jabfm.17.6.424
Min, H.-K., Kapoor, A., Fuchs, M., Mirshahi, F., Zhou, H., Maher, J., et al. (2012). Increased hepatic synthesis and dysregulation of cholesterol metabolism is associated with the severity of nonalcoholic fatty liver disease. *Cell Metab.* 15, 665–674. doi: 10.1016/j.cmet.2012.04.004

Moore, K. J., Rayner, K. J., Suárez, Y., and Fernández-Hernando, C. (2010). miRNAs and cholesterol metabolism. *Trends Endocrinol. Metab.* 21, 699–706. doi: 10.1016/j.tem.2010.08.008

Murray, B. S., Choe, S. E., Woods, M., Ryan, T. E., and Liu, W. (2010). An *in silico* analysis of microRNAs: mining the miRNAome. *Mol. Biosyst.* 6, 1853–1862. doi: 10.1039/c003961f

Naeem, H. K., Robert, C., Gergely, Z., and Zimmer, R. (2010). miRSel: automated extraction of associations between microRNAs and genes from the biomedical literature. *BMC Bioinformatics* 11:135. doi: 10.1186/1471-2105-11-135

Najafi-Shoushtari, S. H., Kristo, F., Li, Y., Shiota, T., Cohen, D. E., Gerszten, R. E., et al. (2010). MicroRNA-33 and the SREBP host genes cooperate to control cholesterol homeostasis. *Science* 328, 1566–1569. doi: 10.1126/science.1189123

Osborne, T. F., and Espeneshade, P. J. (2009). Evolutionary conservation and adaptation in the mechanism that regulates SREBP action: what a long, strange tRIP it’s been. *Genes Dev.* 23, 2578–2591. doi: 10.1101/gad.1854309

Rader, D. J., Alexander, E. T., Weibel, G. L., Billheimer, J., and Rothblat, G. H. (2009). The role of reverse cholesterol transport in animals and humans and relationship to atherosclerosis. *J. Lipid Res.* 50, S189–S194. doi: 10.1194/jlr.R800088-JLR200

Rayner, K. J., Suárez, Y., Dávalos, A., Parathath, S., Fitzgerald, M. L., Tamehiro, N., et al. (2010). MiR-33 contributes to the regulation of cholesterol homeostasis. *Science* 328, 1570–1573. doi: 10.1126/science.1189862

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504. doi: 10.1101/gr.1239303

Tall, A. R., Yvan-Charvet, L., Terasaka, N., Pagler, T., and Wang, N. (2008). HDL, ABC transporters, and cholesterol efflux: implications for the treatment of atherosclerosis. *Cell Metab.* 7, 365–375. doi: 10.1016/j.cmet.2008.03.001

Vickers, K. C., Shoucri, B. M., Levin, M. G., Wu, H., Pearson, D. S., Osei-Hwedieh, D., et al. (2013). MicroRNA-27b is a regulatory hub in lipid metabolism and is altered in dyslipidemia. *Hepatology* 57, 533–542. doi: 10.1002/hep.25846

Wang, X., and Rader, D. J. (2007). Molecular regulation of macrophage reverse cholesterol transport. *Curr. Opin. Cardiol.* 22, 368–372. doi: 10.1097/HCO.0b013e3281ec5113

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Copyright © 2017 Kandhro, Shoombuatong, Nantasenamat, Prachayasittikul and Nuchnoi. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.**