A peer-reviewed version of this preprint was published in PeerJ on 6 August 2019.

View the peer-reviewed version (peerj.com/articles/7309), which is the preferred citable publication unless you specifically need to cite this preprint.

Liu H, Peng Y, Lee H. 2019. miRDRN—miRNA disease regulatory network: a tool for exploring disease and tissue-specific microRNA regulatory networks. PeerJ 7:e7309 https://doi.org/10.7717/peerj.7309

miRDRN – miRNA Disease Regulatory Network: A tool for exploring disease and tissue-specific microRNA regulatory networks

Hsueh-Chuan Liu [1], **Yi-Shian Peng** [1], **Hoong-Chien Lee** [Corresp. 1]

[1] Department of Biomedical Sciences and Engineering, National Central University, Taoyuan City, Taiwan

Corresponding Author: Hoong-Chien Lee
Email address: hclee12345@gmail.com

Background. MiRNA regulates cellular processes through acting on specific target genes. Hundreds of miRNAs and their target genes have been identified, as are many miRNA-disease associations. Cellular processes, including those related to disease, proceed through multiple interactions, are often organized into pathways among genes and gene products. Large databases on protein-protein interactions (PPIs) are available. Here, we have integrated the information mentioned above to build a web service platform, miRNA Disease Regulatory Network, or miRDRN, for users to construct disease and tissue-specific miRNA-protein regulatory networks. **Methods.** Data on human protein interaction, disease-associated miRNA, tumor-associated gene, miRNA targeted gene, molecular interaction and reaction network or pathway, gene ontology, gene annotation and gene product information, and gene expression were collected from publicly available databases and integrated. A complete set of regulatory sub-pathways (RSPs) having the form (M, T, G_1, G_2) were built from the integrated data and stored in the database part of miRDRN, where M is a disease-associated miRNA, T is its regulatory target gene, G_1 (G_2) is a gene/protein interacting with T (G_1). Each sequence (T, G_1, G_2) was assigned a p-value weighted by the participation of the three genes in molecular interactions and reaction pathways. **Results.** A web service platform, miRDRN (http://mirdrn.ncu.edu.tw/mirdrn/), was built to allow users to retrieve a disease and tissue-specific subset of RSPs, from which a miRNA regulatory network is constructed. miRDRN is a database that currently contains 6,973,875 p-valued sub-pathways associated with 119 diseases in 78 tissue types built from 207 diseases-associated miRNA regulating 389 genes, and a web tool that facilitates the construction and visualization of disease and tissue-specific miRNA-protein regulatory networks, for exploring single diseases, or for exploring the comorbidity of disease-pairs. As demonstrations, miRDRN was applied: to explore the single disease colorectal cancer (CRC), in which 26 novel potential CRC target genes were identified; to study the

comorbidity of the disease-pair Alzheimer's disease-Type 2 diabetes (AD-T2D), in which 18 novel potential comorbid genes were identified; and, to explore possible causes that may shed light on recent failures of late-phase trials of anti-AD, *BACE1* inhibitor drugs, in which genes downstream to *BACE1* whose suppression may affect signal transduction were identified.

miRDRN – miRNA Disease Regulatory Network: A tool for exploring disease and tissue-specific microRNA regulatory networks

Hsueh-Chuan Liu[1], Yi-Shian Peng[1] and Hoong-Chien Lee.[1,2]

[1]Department of Biomedical Sciences and Engineering, National Central University, Zhongli District, Taoyuan City, 32001, Taiwan

[2]Department of Physics, Chung Yuan Christian University, Zhongli District, Taoyuan City, 32023, Taiwan

Corresponding Author:
Hoong-Chien Lee
300 Zhongda Road, Zhongli District Taoyuan City, 32001, Taiwan
Email address: hclee12345@gmail.com

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27699v1 | CC BY 4.0 Open Access | rec: 2 May 2019, publ: 2 May 2019

Abstract

Background. MiRNA regulates cellular processes through acting on specific target genes. Hundreds of miRNAs and their target genes have been identified, as are many miRNA-disease associations. Cellular processes, including those related to disease, proceed through multiple interactions, are often organized into pathways among genes and gene products. Large databases on protein-protein interactions (PPIs) are available. Here, we have integrated the information mentioned above to build a web service platform, miRNA Disease Regulatory Network, or miRDRN, for users to construct disease and tissue-specific miRNA-protein regulatory networks.
Methods. Data on human protein interaction, disease-associated miRNA, tumor-associated gene, miRNA targeted gene, molecular interaction and reaction network or pathway, gene ontology, gene annotation and gene product information, and gene expression were collected from publicly available databases and integrated. A complete set of regulatory sub-pathways (RSPs) having the form (M, T, G_1, G_2) were built from the integrated data and stored in the database part of miRDRN, where M is a disease-associated miRNA, T is its regulatory target gene, G_1 (G_2) is a gene/protein interacting with T (G_1). Each sequence (T, G_1, G_2) was assigned a p-value weighted by the participation of the three genes in molecular interactions and reaction pathways.
Results. A web service platform, miRDRN (http://mirdrn.ncu.edu.tw/mirdrn/), was built to allow users to retrieve a disease and tissue-specific subset of RSPs, from which a miRNA regulatory network is constructed. miRDRN is a database that currently contains 6,973,875 p-valued sub-pathways associated with 119 diseases in 78 tissue types built from 207 diseases-associated miRNA regulating 389 genes, and a web tool that facilitates the construction and visualization of disease and tissue-specific miRNA-protein regulatory networks, for exploring single diseases, or for exploring the comorbidity of disease-pairs. As demonstrations, miRDRN was applied: to explore the single disease colorectal cancer (CRC), in which 26 novel potential CRC target genes were identified; to study the comorbidity of the disease-pair Alzheimer's disease-Type 2 diabetes (AD-T2D), in which 18 novel potential comorbid genes were identified; and, to explore possible causes that may shed light on recent failures of late-phase trials of anti-AD, *BACE1* inhibitor drugs, in which genes downstream to *BACE1* whose suppression may affect signal transduction were identified.

Keywords: Diseases, database, service tool, disease-associate miRNA, disease and tissue-specific miRNA-protein regulatory pathway, disease target gene, comorbidity gene, colorectal cancer, Alzheimer's disease, Type 2 diabetes, anti-AD *BACE1* inhibitor drug

Introduction

Protein-protein interactions (PPIs) are critical to almost all biological process, and a good knowledge of the network of interacting proteins is crucial to understanding cellular mechanisms [1]. Recent advances in biotechnology, such as high-throughput yeast two-hybrid screening, have allowed scientists to build maps of proteome-wide PPI, or interactome. Conventionally, a PPI map is a static network, in which each node represents a protein and an edge connecting two proteins indicates that there is experimental evidence showing that, under certain circumstances, the two proteins would interact. In reality, a PPI network (PPIN) should be viewed as a dynamic entity: it is an interaction network that is intrinsically controlled by regulatory mechanisms and changes with time and space [2], as determined by the physiological condition of the cell in which the proteins reside. If there is a PPIN that includes all possible PPIs, then, under a specific physiological condition only a specific sub-network of the PPIN is realized.

MicroRNAs (miRNAs) are small (~22 nucleotides) noncoding regulatory RNA molecules in plants, animals, and some viruses. In a process known as RNA interference (RNAi), a miRNA regulates gene expression by destabilizing and/or disrupting the translation of fully or partially sequenced mRNA [3, 4]. In this way a miRNA regulates the formation of all PPINs to which its target is connected, and by extension all biological processes with which those PPINs are involved. As well as acting as a tumor suppressor gene, a miRNA may also act as an oncogene, say, by targeting a tumor suppressor gene [5]. The function of a specific biological process, or its malfunction, such as associated with a disease, typically involves a complex composed of a set of miRNA-regulated proteins, together with their interacting protein partners. The study of such miRNA-protein complexes should be an integral part of understanding biological processes [6] as well as diseases.

An understanding of the molecular and physio-pathological mechanisms of diseases is crucial for the design of disease preventive and therapeutic strategies. The combination of experimental and computational methods has led to the discovery of disease-related genes [7, 8]. An example is the causal relation connecting the malfunction causing mutations in the enzyme phenylalanine hydroxylase to the metabolic disorder Phenylketonuria [9]. Many human diseases cannot be attributed to single-gene malfunctions but arise from complex interactions among multiple genetic variants [10]. How a disease is caused and how it can be treated can be better studied on the basis of a body of knowledge including all associated genes and biological pathways involving those genes.

Diseases are usually defined by a set of phenotypes that are associated with various pathological processes and their mutual interactions. Some relations between phenotypes of different diseases

may be understood on the basis of common underlying molecular processes [11], such as when there are genes associated to both diseases. It has been shown that genes associated with the same disorder encode proteins that have a strong tendency to interact with each other [12]. More specifically, one may consider two diseases to be related if their metabolic reactions within a cell share common enzymes [13]. Networks of PPIs have also been studied in the context of disease interactions [14, 15].

Here, we report on a web service platform, miRNA Disease Regulatory Network (miRDRN). The platform contains two parts, a database that, in its current form, contains 6,973,875 *p*-valued miRNA regulatory sub-pathways associated with 119 diseases in 78 tissue types built from 207 diseases-associated miRNA regulating 389 genes; and a novel web-based tool that, using the miRDRN database and public protein-protein database, facilitates the construction and visualization of miRNA regulatory networks for user specified single diseases and, for comorbidity studies, disease-pairs. We demonstrate three applications of miRDRN: to explore the molecular and network properties of the single disease colorectal neoplasm; to study the comorbidity of the disease-pair, Alzheimer's disease (AD) and Type 2 diabetes; and, by using miRDRN to construct a miRNA regulatory sub-network centered on the gene *BACE1*, to look for insights that may explain why several anti-AD, *BACE1* inhibiting drugs that failed recent late-phase trials worsened conditions of treatment groups.

Materials and Methods

Data integration

Data on human protein interaction (BioGRID [16]), disease-associated miRNA (HMDD [17]), tumor-associated gene (TAG [18]), miRNA targeted gene (HMDD, TarBase [19]), molecular interaction and reaction network or pathway (KEGG [20]), gene ontology, gene annotation and gene product information (GO), and gene expression (GeneBank) were collected from publicly available data bases (Table 1) and integrated.

Construction of miRNA regulatory sub-pathways

We define a regulatory sub-pathway (RSP) as a linked sequence (M, T, G_1, G_2) (Figure 1), where M is a miRNA, T is its regulatory target gene [18,19,20], G_1 is a gene whose encoded protein (p_1) interacts (according to PPI data) with the protein (p_T) encoded by T, and G_2 is a gene whose encoded protein (p_2) interacts with p_1. In what follows, when there is little risk of misunderstanding, the same symbol will be used to represent a gene or the protein it encodes. The idea of RSP construction is this: given a miRNA and a target gene T, we use PPI data [17] to

collect all RSPs by extending from T two levels of interaction. For a given disease, all such sub-pathways emanating from every one of the known miRNAs associated with the disease [18] were constructed.

Jaccard score of a regulatory sub-pathway

The Jaccard similarity coefficient base [23] were used to score the RSPs, based on the assumption that there is a tendency for two directly interacting proteins to participate in the same set of biological processes or share the same set of molecular functions. Given two sets $S1$ and $S2$ (in the current application, a set will be either a list of biological processes (BP) or a list of molecular functions (MF), both according to GO [21]), the Jaccard coefficient (JC) of $S1$ and $S2$ is defined as,

$$JC\,(S1,\,S2) = \frac{|S1 \cap S2|}{|S1 \cup S2|}$$

Where $\cap$ is the union (of two sets), $\cup$ is the intersection, and $|Z|$ is the cardinality of Z. JC, which ranges from 0 to 1, is a quantitative measure of the similarity between two sets. For example, when $S1 = \{a, b, c\}$ and $S2 = \{b, c, d\}$, $JC\,(S1, S2) = 2/4 = 0.5$.

Let (M, T, G_1, G_2) be an RSP as defined in the previous section and denote by $[G]$ the set of biological processes (or pathways) [20,21] that involve the gene G. We define the Jaccard score, or JS, of RSP as,

$$JS_X\,(T,\,G_1,\,G_2) = \frac{1}{2}(JC\,([T]_X,\,[G_1]_X) + JC\,([G_1]_X,\,[G_2]_X))$$

Where X may be BP or MF. If the pair $[T]$ and $[G_1]$ do not share a common term, then the corresponding JC has a zero value; similarly for the pair $[G_1]$ and $[G_2]$. In either case the RSP, (M, T, G_1, G_2), whatever M is, is considered to be not viable and discarded. Note that the JS of an RSP depends only on the genes in the pathway, not on the miRNA. There could be multiple RSPs emanating from a miRNA associated with a disease, and these RSPs may be ranked by their JS's.

***P*-value of a sub-pathway**

A p-value for an RSP (M, T, G_1, G_2), independent of M, was assigned as follows. Let the total number of BP (or MF, as the case may be) terms be N, and the number of terms in $[T]$, $[G_1]$, $[G_2]$, $[T] \cap [G_1]$, $[G_1] \cap [G_2]$ be x, y, z, n_1, and n_2, respectively, then the p-values, P_1 and P_2, for (T, G_1) and (G_1, G_2) are respectively

$$P_1 = \frac{C_{n_1}^{N} C_{x-n_1}^{N-n_1} C_{y-n_1}^{N-x}}{C_x^N C_y^N}$$

and

$$P_2 = \frac{C_{n_2}^{N} C_{y-n_2}^{N-n_2} C_{z-n_2}^{N-y}}{C_y^N C_z^N}$$

The *p*-value for the RSP was set to be the greater of P_1 and P_2.

Construction of disease-associated miRNA regulatory network

A disease-associated miRNA regulatory network (RRN) is constructed as follows. Step 1. Select a disease. Step 2. Collect all miRNAs (*M*'s) associated with the disease from HMDD [17]. Step 3. Collect all target genes (*T*'s) of the collected miRNAs from HMDD [17] and TarBase [19]. Step 4. Construct all RSPs (having the from (M, T, G_1, G_2)) using PPI data (BioGRID) [16] and compute the Jaccard coefficients (JCs) of the two PPIs in each of the RSPs. If either one of the JCs has zero value discard the RSP, otherwise the Jaccard score of the RSP is taken to be the mean of the two JCs. Compute the *p*-values of the RSPs. Step 5. Construct an RRN from entire set of generated RSPs by linking (pairs of) genes from different RSPs whenever the proteins coded by the genes have interaction according to PPI data (Figure 2).

Results and Discussion

miRNA Disease Regulatory Network (miRDRN) – A web service platform

We built miRDRN (http://mirdrn.ncu.edu.tw/mirdrn/), a web-based service that allows the user to construct a disease and tissue-specific, *p*-valued, miRNA regulatory gene network, or miRNA regulatory network (RRN). The current version of miRDRN contains 6,973,875 *p*-valued RSPs constructed through 389 miRNA-regulated genes from 207 diseases-associated miRNAs associated with 119 diseases (Table 2).

User may use miRDRN to explore a single disease, or the comorbidity of a disease-pair. In the course of either type of study, all relevant miRNAs, genes, and RSPs are made accessible to the user in tabulated form, and RRNs in the form of interactive maps, both of which may be

downloaded by the user. Often a map is too large for practical visualization, and in such a case the user may use options such as setting a *p*-value cut-off, or requiring a specific gene to be present in the map, or both, to obtain a partial RRN.

The entrance interface of miRDRN (http://mirdrn.ncu.edu.tw/mirdrn/) asks the user to select "Single Search" to explore a single disease or "Comorbidity Search" to explore the comorbidity of a disease-pair (Figure 3). Next the user is asked to specify the disease or disease-pair to be explored and tissue/tumor types, and *p*-value threshold for RSP evaluation, as well as several optional inputs. The user may then click on "Query" to start the start the search engine (Figure 4). Tabulated results of diseaseassociated miRNAs and their target genes (Figure 5), a multi-page list of all RSPs (Figure 6) and, in the case of Comorbidity Search, a list of all comorbid genes (Figure 7) will then automatically appear. After the first, automatic iteration, the user may reduce the size of the RSP-list by using the "Gene filter" and "Show top ... sub-pathways" options (Figure 6). The next interface (Figure 8), in ready mode on first appearance, waits for the user to select one of three network layouts: "Tree", "Circle", or "Radial". After "Go" is clicked on, the platform displays an interactive map showing the RRN built from RSPs selected by user-specified options (Figure 8). When the mouse is placed on a node (a miRNA or a gene) on the map a small text window opens to show the name of the node and annotations from GO, OMIM, KEGG and GeneBank databases.

Three applications of miRDRN

Case 1. A single disease study of colorectal neoplasm

Here we demonstrate a single disease application of miRDRN. After logging onto miRDRN's main interface (Figure 3), click on "Single Search" to see a new window and select a disease and other options as desired. For the present case "colorectal neoplasms" (or colorectal cancer, CRC), tissue type "colorectal tumor", pathway ranking by "Jaccard index (MF)", and p-value < "0.001" were selected. Then click on "Query" to start. The query yielded 33 associated-miRNAs, 37 miRNA regulated genes, and 45,565 RSPs involving 3,079 genes (Table 3).

By default, the interface "miRNA regulatory sub-pathways" (Figure 6) lists all the constructed RSPs, namely all 45,565 of them in the present case and, if requested, would present a drawing including all the RSPs which, however, would be difficult to visualize, not to say interact with. On the same interface are two options for displaying/using a smaller RSP set: "Gene filter", where the user can restrict the set to only those RSPs containing a specified gene; and "Show top … sub-pathways", where the user can ask for only the N-top

RSPs having the smallest *p*-values be listed and used for network construction. The interface "Disease specific miRNA regulatory network" then allows the user to choose one among the layouts "Tree", "Circle", and "Radial". Here a tree-map, with several disconnected parts, built from the top-70 RSPs is shown (Figure 9).

In computer mode, when the mouse is placed on a node in the network, a small text window opens showing the name of the node/gene and its weight, or the number of other nodes it is connected to in the 45,565-node network. Its largest connected sub-RRN, or "Network-1" (Figure 10), is composed of six miRNAs targeting four genes connected to 52 other genes, (Table 4). Of the 56 genes in Network-1, 22 have known CRC connections (CORECG database, http://lms.snu.edu.in/corecg) [24], and 26 other have references linking them either directly or indirectly to CRC [25-52] (Table 5). Among these, *TNIK* [31] and *TNK2* [51] have been used as drug targets for CRC treatment. We consider the remaining eight genes - *PRKACA, MAP3K12, LRRK1, RIOK2, OXSR1, CDK17, EIF2AK1, TSSK4* – to be potential, novel CRC-related genes. Noticeably, Network-1 has two parts, one 28 nodes (five miRNAs targeting three genes) and the other 34 nodes (one miRNA targeting one gene), connected by a single link, or PPI. The three types of genes, known CRC-related, reference-supported, and potential CRC-related, are more or less proportionately distributed in these two parts.

The "Gene filter" option (Figure 6) allows the user to focus on a specific gene in RRN construction. As example, *TNK2*, a key drug target for the treatment of metastatic CRC [51], was selected as the filter, together with the "Show top 70 RSPs" option. The result was a nine-node sub-RRN: the target gene *AXL* regulated by three miRNAs – hsa-mir-199b, hsa-mir-34a, hsa-mir-199a – and linked (by PPI) to *TNK2*, itself linked to four other genes *AXL*(OCG)*, MAGI3, HSP90AB2P, MERTK*(OCG)*, KAT8* (Figure 11).

Case 2. A Comorbidity study of AD and T2D

Here we demonstrate a two-disease application of miRDRN. After logging onto miRDRN's main interface (Figure 3), click on "Comorbidity Search" to see a new window urging the user to select two diseases; for "Disease 1", "Alzheimer Disease" (AD) and tissue type "brain" were selected and for "Disease 2", "Type 2", which stands for type 2 diabetes (T2D), and tissue type "pancreas". Pathway ranking by "Jaccard index (MF)", and *p*-value < "0.005" for both diseases were selected. Both AD and T2D are complex diseases and share aging for a risk factor; accumulated evidence indicates a connection between these two diseases at the molecular level [53]. For this case the program yielded, for AD (T2D), three (one)

associated-miRNAs, three (one) targeted genes, 644 (3908) RSPs, involving 633 (2187) genes (Table 6).

Because the two did not have any common associated-miRNA target gene, they had distinct sets of RSPs. However, with 500 genes, call "comorbid" genes, in the two sets of RSPs being common, the two sets of RSPs had only 2320 distinct genes (Table 6). Among the comorbid genes, 8 - *ALOX5, APP, BIN1, CHGB, VWF, NEFL, LETMD1, CELF1*- were identified as known AD target genes [56, 57, 58] and 14 - *TCF7L2, APOA1, VWF, CDKN2B, CAT, ITGB2, ISL1, POLD3, APP, NFKBIB, GNA12, DEDD, LDLR, PRKAB1*- as known T2D target genes [59], *APP* and *VWF* are known targets of both diseases (Table 7). With the exception of three - *LEMD1*, *POLD3, GNA12*, the comorbidity of all the others have literature support (Table 7).

Case 3. A sub-RRN centered on the AD-associated gene *BACE1*

In recent years a number of anti-AD drugs designed on the basis of the amyloid-beta (Aβ) hypothesis of AD, which holds that Aβ aggregate in the brain is the main causative factor of AD, failed late-phase trials. These include the γ-secretase inhibitor Semagacestat [89] and two *BACE1* inhibitors, Verubecestat [90] and Atabecestat [91]. In all three cases treatment groups scored worse than the control group on the ADCS-ADL (Alzheimer's Disease Cooperative Study Activities of Daily Living Inventory) functional measure and reported more anxiety, depression, and sleep problems than controls. In a "Single Search" application on AD (tissue, brain; *p*-value threshold, 0.005), we had miRDRN construct a partial RRN (Gene filter, *BACE1*; Show top 70 sub-pathways; Network layout, Radial) centered on *BACE1*, which is a regulatory target of hsa-mir-195. The result shows the genes *PSEN1*, *NCSTN*, *RANBP9*, *PLSCR1*, *MMP2*, and *FURIN* to be immediately downstream to *BACE1* in the RRN (Figure 12). *PSEN1* and *NCSTN* encode proteins that are, respectively, catalytic and essential subunits of the γ-secretase complex; suppression of these genes are presumably the purpose of *BACE1* inhibition. On the other hand, *RANBP9* encodes a protein that facilitates the progression of mitosis in developing neuroepithelial cells [92]; *PLSCR1* encodes a protein that acts in the control of intracellular calcium homeostasis and has a central role in signal transduction [93]; *MMP2* encodes a protein that promotes neural progenitor cell migration [94]. Suppression of these genes (by *BACE1* inhibition) may therefore adversely affect signal transduction and the nerve system, and could be part of the reason why Semagacestat, Verubecestat, and Atabecestat worsened the ADCS-ADL functional measure of treatment groups.

Conclusion

This work describes a web service platform, miRDRN, composed of a new database and a web-based tool, for constructing miRNA regulatory networks for the user to explore the molecular and regulatory network properties of single diseases as well as for pairs of diseases. As demonstration, miRDRN was applied to study the single disease CRC, where 34 potential target genes were identified, 26 of which have literature support; to study the comorbidity of the disease-pair AD-T2D, where 20 potential novel AD-T2D comorbid genes were identified, 17 of which have literature support; and to construct a partial miRNA regulatory sub-network centered on the AD-associated gene *BACE1*, which in turn suggests a possible explanation why, in late-phase trials that ended in failure, several γ/β-secretase inhibiting anti-AD drugs worsened the functional measure of treatment groups. We believe miRDRN is a useful tool for exploring the molecular and network properties of single diseases and those connecting pairs of diseases, and for discovering new insights on the molecular properties, including potential side effects, of disease treating drugs.

Abbreviations

AD: Alzheimer's Disease
CRG: Cancer related gene
GO: Gene Ontology
KEGG: The Kyoto Encyclopedia of Genes and Genomes
miRDRN: The miRNA Disease Regulatory network web service platform
OCG: Oncogene
PPI: protein-protein interaction
PPIN: PPI network
RSP: Regulatory sub-network
RRN: Disease-associated miRNA regulatory network
T2D: Type 2 Diabetes
TSG: Tumor suppressor gene

References

1. Rual JF, Venkatesan K, Hao T, Hirozane-Kishikawa T, Dricot A, Li N, Berriz G, Gibbons G, Dreze M, Ayivi-Guedehoussou N, et al. Towards a proteome-scale map of the human protein-protein interaction network. Nature. 2005; 437: 1173-1178.
2. LIANG H, LI WH. MicroRNA regulation of human protein-protein interaction network. RNA. 2007; 13: 1402-1408.
3. Bartel DP. MicroRNAs: target recognition and regulatory functions. Cell. 2009; 136(2): 215–233.
4. Landgraf P, Rusu M, Sheridan R, Sewer A, Iovino N, Aravin A, Pfeffer S, Rice A, Kamphorst AO, Landthaler M, et al. A mammalian microRNA expression atlas based on small RNA library sequencing. Cell. 2007; 129(7): 1401-14.
5. Zhang W, Dahlberg JE, Tam W. MicroRNAs in tumorigenesis: a primer. Am J Pathol. 2007; 171(3): 728-38.
6. Hsu CW, Juan HF, Huang HC. Characterization of microRNA-regulated protein-protein interaction network. Proteomics. 2008; 8(10): 1975-9.
7. Botstein D, Risch N. Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. Nature Genetics. 2003; 33: 228–237.
8. Kann MG. Advances in translational bioinformatics: computational approaches for the hunting of disease genes. Brief Bioinform. 2010; 11: 96–110.
9. Scriver CR, Waters PJ. Monogenic traits are not simple: lessons from phenylketonuria. Trends Genet. 1999; 15: 267–272.
10. Hirschhorn JN, Daly MJ. Genome-wide association studies for common diseases and complex traits. Nat Rev Genet. 2005; 6: 95–108.
11. Barabási AL, Gulbahce N, Loscalzo J. Network medicine: a network-based approach to human disease. Nat Rev. Genet. 201; 12(1): 56-68.
12. Goh KI, Cusick ME, Valle D, Childs B, Vidal M, Barabási AL. The human disease network. Proc Natl Acad Sci U S A. 2007; 104(21): 8685-90.
13. Lee DS, Park J, Kay KA, Christakis NA, Oltvai ZN, Barabási AL. The implications of human metabolic network topology for disease comorbidity. Proc Natl Acad Sci U S A. 2008; 105(29): 9880-5.
14. Ideker T, Sharan R. Protein networks in disease. Genome Res. 2008; 18(4): 644-52.
15. Lim J, Hao T, Shaw C, Patel AJ, Szabó G, Rual JF, Fisk CJ, Li N, Smolyar A, Hill DE, et al. A protein-protein interaction network for human inherited ataxias and disorders of Purkinje cell degeneration. Cell. 2006; 125(4): 801-14.

16. Chatr-Aryamontri A, Breitkreutz BJ, Heinicke S, Boucher L, Winter A, Stark C, Nixon J, Ramage L, Kolas N, O'Donnell L, et al. The BioGRID interaction database: 2013 update. Nucleic Acids Res. 2013; 41(Database issue): D816-23.
17. Li Y, Qiu C, Tu J, Geng B, Yang J, Jiang T, Cui Q. HMDD v2.0: a database for experimentally supported human microRNA and disease associations. Nucleic Acids Res. 2014; 42(Database issue): D1070-4.
18. Chen JS, Hung WS, Chan HH, Tsai SJ, Sun HS. In silico identification of oncogenic potential of fyn-related kinase in hepatocellular carcinoma. Bioinformatics. 2013; 29(4): 420-7.
19. Papadopoulos GL, Reczko M, Simossis VA, Sethupathy P, Hatzigeorgiou AG. The database of experimentally supported targets: a functional update of TarBase. Nucleic Acids Res. 2009; 37(Database issue): D155-8.
20. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Res. 2017; 45(D1): D353-D361.
21. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 2000 May;25(1):25-9.
22. Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. GenBank. Nucleic Acids Res. 2013 Jan;41(Database issue):D36-42.
23. Ng KL, Liu HC, Lee SC. ncRNAppi-a tool for identifying disease-related miRNA and siRNA targeting pathways. Bioinformatics. 2009; 25(23): 3199-201.
24. Agarwal R, Kumar B, Jayadev M, Raghav D, Singh A. CoReCG: a comprehensive database of genes associated with colon-rectal cancer. Database (Oxford). 2016 Apr 25;2016.
25. Kang MH, Moon SU, Sung JH, Kim JW, Lee KW, Lee HS, Lee JS, Kim JH. Antitumor Activity of HM781-36B, alone or in Combination with Chemotherapeutic Agents, in Colorectal Cancer Cells. Cancer Res Treat. 2016 Jan;48(1):355-64.
26. Sirvent A, Bénistant C, Pannequin J, Veracini L, Simon V, Bourgaux JF, Hollande F, Cruzalegui F, Roche S. Src family tyrosine kinases-driven colon cancer cell invasion is induced by Csk membrane delocalization. Oncogene. 2010 Mar 4;29(9):1303-15.
27. Jeong D, Kim H, Kim D, Ban S, Oh S, Ji S, Kang D, Lee H, Ahn TS, Kim HJ, Bae SB, Lee MS, Kim CJ, Kwon HY, Baek MJ. Protein kinase, membrane-associated tyrosine/threonine 1 is associated with the progression of colorectal cancer. Oncol Rep. 2018 Jun;39(6):2829-2836.
28. Xie T, D' Ario G, Lamb JR, Martin E, Wang K, Tejpar S, Delorenzi M, Bosman FT, Roth AD, Yan P, Bougel S, Di Narzo AF, Popovici V, Budinská E, Mao M, Weinrich SL, Rejto

PA, Hodgson JG. A comprehensive characterization of genome-wide copy number aberrations in colorectal cancer reveals novel oncogenes and patterns of alterations. PLoS One. 2012;7(7):e42001.

29. Wu S, Wu F, Jiang Z. Identification of hub genes, key miRNAs and potential molecular mechanisms of colorectal cancer. Oncol Rep. 2017 Oct;38(4):2043-2050.
30. Xiang Z, Wang S, Xiang Y. Up-regulated microRNA499a by hepatitis B virus induced hepatocellular carcinogenesis via targeting MAPK6. PLoS One. 2014 Oct 23;9(10):e111410.
31. Masuda M, Yamada T. The emergence of TNIK as a therapeutic target for colorectal cancer. Expert Opin Ther Targets. 2017 Apr;21(4):353-355.
32. Ali RH, Marafie MJ, Bitar MS, Al-Dousari F, Ismael S, Bin Haider H, Al-Ali W, Jacob SP, Al-Mulla F. Gender-associated genomic differences in colorectal cancer: clinical insight from feminization of male cancer cells. Int J Mol Sci. 2014 Sep 29;15(10):17344-65.
33. Yun CW, Kim S, Lee JH, Lee SH. Melatonin Promotes Apoptosis of Colorectal Cancer Cells via Superoxide-mediated ER Stress by Inhibiting Cellular Prion Protein Expression. Anticancer Res. 2018 Jul;38(7):3951-3960.
34. Vázquez-Cedeira M, Lazo PA. Human VRK2 (vaccinia-related kinase 2) modulates tumor cell invasion by hyperactivation of NFAT1 and expression of cyclooxygenase-2. J Biol Chem. 2012 Dec 14;287(51):42739-50.
35. Zhang YJ, Dai Q, Sun DF, Xiong H, Tian XQ, Gao FH, Xu MH, Chen GQ, Han ZG, Fang JY. mTOR signaling pathway is a target for the treatment of colorectal cancer. Send to Ann Surg Oncol. 2009 Sep;16(9):2617-28.
36. Csukasi F, Duran I, Barad M, Barta T, Gudernova I, Trantirek L, Martin JH, Kuo CY, Woods J, Lee H, Cohn DH, Krejci P, Krakow D. The PTH/PTHrP-SIK3 pathway affects skeletogenesis through altered mTOR signaling. Sci Transl Med. 2018 Sep 19;10(459). pii: eaat9356.
37. Rey C, Faustin B, Mahouche I, Ruggieri R, Brulard C, Ichas F, Soubeyran I, Lartigue L, De Giorgi F. The MAP3K ZAK, a novel modulator of ERK-dependent migration, is upregulated in colorectal cancer. Oncogene. 2016 Jun 16;35(24):3190-200.
38. Goyal P, Behring A, Kumar A, Siess W. Identifying and characterizing a novel protein kinase STK35L1 and deciphering its orthologs and close-homologs in vertebrates. PLoS One. 2009 Sep 16;4(9):e6981.
39. Sabir SR, Sahota NK, Jones GD, Fry AM. Loss of Nek11 Prevents G2/M Arrest and Promotes Cell Death in HCT116 Colorectal Cancer Cells Exposed to Therapeutic DNA Damaging Agents. PLoS One. 2015 Oct 26;10(10):e0140975.
40. Bjerrum JT, Nielsen OH, Riis LB, Pittet V, Mueller C, Rogler G, Olsen J. Transcriptional analysis of left-sided colitis, pancolitis, and ulcerative colitis-associated dysplasia. Inflamm

Bowel Dis. 2014 Dec;20(12):2340-52.
41. Hanna DL, Loupakis F, Yang D, Cremolini C, Schirripa M, Li M, Matsusaka S, Berger MD, Miyamoto Y, Zhang W, Ning Y, Antoniotti C, Salvatore L, Moran M, Zeger G, Astrow SH, Falcone A, Lenz HJ. Prognostic Value of ACVRL1 Expression in Metastatic Colorectal Cancer Patients Receiving First-line Chemotherapy With Bevacizumab: Results From the Triplet Plus Bevacizumab (TRIBE) Study. Clin Colorectal Cancer. 2018 Sep;17(3):e471-e488.
42. Record CJ, Chaikuad A, Rellos P, Das S, Pike AC, Fedorov O, Marsden BD, Knapp S, Lee WH. Structural comparison of human mammalian ste20-like kinases. PLoS One. 2010 Aug 6;5(8):e11905.
43. Zhou JK, Zheng YZ, Liu XS, Gou Q, Ma R, Guo CL, Croce CM, Liu L, Peng Y. ROR1 expression as a biomarker for predicting prognosis in patients with colorectal cancer. Oncotarget. 2017 May 16;8(20):32864-32872.
44. Gong H, Fang L, Li Y, Du J, Zhou B, Wang X, Zhou H, Gao L, Wang K, Zhang J. miR-873 inhibits colorectal cancer cell proliferation by targeting TRAF5 and TAB1. Oncol Rep. 2018 Mar;39(3):1090-1098.
45. Yuan WC, Pepe-Mooney B, Galli GG, Dill MT, Huang HT, Hao M, Wang Y, Liang H, Calogero RA, Camargo FD. NUAK2 is a critical YAP target in liver cancer. Nat Commun. 2018 Nov 16;9(1):4834.
46. Kim ST, Ahn TJ, Lee E, Do IG, Lee SJ, Park SH, Park JO, Park YS, Lim HY, Kang WK, Kim SH, Lee J, Kim HC. Exploratory biomarker analysis for treatment response in KRAS wild type metastatic colorectal cancer patients who received cetuximab plus irinotecan. BMC Cancer. 2015 Oct 20;15:747.
47. Li BQ, Huang T, Zhang J, Zhang N, Huang GH, Liu L, Cai YD. An ensemble prognostic model for colorectal cancer. Send to PLoS One. 2013 May 2;8(5):e63494.
48. Guo H, Hu X, Ge S, Qian G, Zhang J. Regulation of RAP1B by miR-139 suppresses human colorectal carcinoma cell proliferation. Int J Biochem Cell Biol. 2012 Sep;44(9):1465-72.
49. Peng H, Luo J, Hao H, Hu J, Xie SK, Ren D, Rao B. MicroRNA-100 regulates SW620 colorectal cancer cell proliferation and invasion by targeting RAP1B. Oncol Rep. 2014 May;31(5):2055-62.
50. Alonso MH, Aussó S, Lopez-Doriga A, Cordero D, Guinó E, Solé X, Barenys M, de Oca J, Capella G, Salazar R, Sanz-Pamplona R, Moreno V. Comprehensive analysis of copy number aberrations in microsatellite stable colon cancer in view of stromal component. Br J Cancer. 2017 Jul 25;117(3):421-431.
51. Qi L, Ding Y. TNK2 as a key drug target for the treatment of metastatic colorectal cancer. Int J Biol Macromol. 2018 Nov;119:48-52.

462 52. Jin DH, Lee J, Kim KM, Kim S, Kim DH, Park J. Overexpression of MAPK15 in gastric
463 cancer is associated with copy number gain and contributes to the stability of c-Jun.
464 Oncotarget. 2015 Aug 21;6(24):20190-203.
465 53. Ahmed F, Ansari JA, Ansari ZE, Alam Q, Gan SH, Kamal MA, Ahmad E. A molecular
466 bridge: connecting type 2 diabetes and Alzheimer's disease. CNS Neurol Disord Drug
467 Targets. 2014; 13(2): 312-21.
468 54. Bertram L, McQueen MB, Mullin K, Blacker D, Tanzi RE. Systematic meta-analyses of
469 Alzheimer disease genetic association studies: the AlzGene database. Nat Genet. 2007;
470 39(1): 17-23.
471 55. Sun J, Feng X, Liang D, Duan Y, Lei H. Down-regulation of energy metabolism in
472 Alzheimer's disease is a protective response of neurons to the microenvironment. J
473 Alzheimers Dis. 2012; 28(2): 389-402.
474 56. Bai Z, Han G, Xie B, Wang J, Song F, Peng X, Lei H. AlzBase: an integrative database for
475 gene dysregulation in Alzheimer’s disease. Mol Neurobiol. 2016; 53(1):310-319.
476 57. Dai HJ, Wu JC, Tsai RT, Pan WH, Hsu WL. T-HOD: a literature-based candidate gene
477 database for hypertension, obesity and diabetes. Database (Oxford). 2013; 2013: bas061.
478 58. Postula M, Janicki PK, Rosiak M, Eyileten C, Zaremba M, Kaplon-Cieslicka A, Sugino S,
479 Kosior DA, Opolski G, Filipiak KJ, Mirowska-Guzel D. Targeted deep resequencing of
480 ALOX5 and ALOX5AP in patients with diabetes and association of rare variants with
481 leukotriene pathways. Exp Ther Med. 2016; 12(1): 415-421.
482 59. Nejatian N, Penna-Martinez M, Steinhilber D, Badenhoop K. The association between
483 vitamin D and the arachidonate 5-lipoxygenase (ALOX-5) gene polymorphism in type 2
484 diabetes. Diabetologie und Stoffwechsel. 2015; 10-P205.
485 60. Heemskerk MM, Giera M, Bouazzaoui FE, Lips MA, Pijl H, van Dijk KW, van Harmelen
486 V. Increased PUFA Content and 5-Lipoxygenase Pathway Expression Are Associated with
487 Subcutaneous Adipose Tissue Inflammation in Obese Women with Type 2 Diabetes.
488 Nutrients. 2015 Sep 11;7(9):7676-90.
489 61. Greenbaum L, Ravona-Springer R, Lubitz I, Schmeidler J, Cooper I, Sano M, Silverman JM,
490 Heymann A, Beeri MS. Potential contribution of the Alzheimer's disease risk locus BIN1 to
491 episodic memory performance in cognitively normal Type 2 diabetes elderly. Eur
492 Neuropsychopharmacol. 2016; 26(4): 787-95.
493 62. Horn S, Kirkegaard JS, Hoelper S, Seymour PA, Rescan C, Nielsen JH, Madsen OD, Jensen
494 JN, Krüger M, Grønborg M, Ahnfelt-Rønne J. Research Resource: A Dual Proteomic
495 Approach Identifies Regulated Islet Proteins During β-Cell Mass Expansion In Vivo. Mol
496 Endocrinol. 2016; 30(1): 133-43.
497 63. Wu DA, Bu X, Warden CH, Shen DD, Jeng CY, Sheu WH, Fuh MM, Katsuya T, Dzau VJ,

Reaven GM, Lusis AJ, Rotter JI, Chen YD. Quantitative trait locus mapping of human blood pressure to a genetic region at or near the lipoprotein lipase gene locus on chromosome 8p22. J Clin Invest. 1996 May 1;97(9):2111-8.

64. Celikbilek A, Tanik N, Sabah S, Borekci E, Akyol L, Ak H, Adam M, Suher M, Yilmaz N. Elevated neurofilament light chain (NFL) mRNA levels in prediabetic peripheral neuropathy. Mol Biol Rep. 2014 Jun;41(6):4017-22.
65. Verma SK, Deshmukh V, Liu P, Nutter CA, Espejo R, Hung ML, Wang GS, Yeo GW, Kuyumcu-Martinez MN. Reactivation of fetal splicing programs in diabetic hearts is mediated by protein kinase C signaling. J Biol Chem. 2013; 288(49): 35372-86.
66. Belanger K, Nutter CA, Li J, Tasnim S, Liu P, Yu P, Kuyumcu-Martinez MN. CELF1 contributes to aberrant alternative splicing patterns in the type 1 diabetic heart. Biochem Biophys Res Commun. 2018 Sep 18;503(4):3205-3211.
67. Blom ES, Wang Y, Skoglund L, Hansson AC, Ubaldi M, Lourdusamy A, Sommer WH, Mielke M, Hyman BT, Heilig M, et al. Increased mRNA Levels of TCF7L2 and MYC of the Wnt Pathway in Tg-ArcSwe Mice and Alzheimer's Disease Brain. Int J Alzheimers Dis. 2010; 2011:936580.
68. Včelák J, Vejražková D, Vaňková M, Lukášová P, Bradnová O, Hálková T, Bešťák J, Andělová K, Kvasničková H, Hoskovcová P, Vondra K, Vrbíková J, Bendlová B. T2D risk haplotypes of the TCF7L2 gene in the Czech population sample: the association with free fatty acids composition. Physiol Res. 2012;61(3):229-40.
69. Arefin AS, Mathieson L, Johnstone D, Berretta R, Moscato P. Unveiling clusters of RNA transcript pairs associated with markers of Alzheimer's disease progression. PLoS One. 2012;7(9):e45535.
70. Riise J, Plath N, Pakkenberg B, Parachikova A. Aberrant Wnt signaling pathway in medial temporal lobe structures of Alzheimer's disease. J Neural Transm (Vienna). 2015 Sep;122(9):1303-18.
71. Vollbach H, Heun R, Morris CM, Edwardson JA, McKeith IG, Jessen F, Schulz A, Maier W, Kölsch H. APOA1 polymorphism influences risk for early-onset nonfamiliar AD. Ann Neurol. 2005 Sep;58(3):436-41.
72. Raygani AV, Rahimi Z, Kharazi H, Tavilani H, Pourmotabbed T. Association between apolipoprotein E polymorphism and serum lipid and apolipoprotein levels with Alzheimer's disease. Neurosci Lett. 2006 Nov 6;408(1):68-72.
73. Lin Q, Cao Y, Gao J. Decreased expression of the APOA1-APOC3-APOA4 gene cluster is associated with risk of Alzheimer's disease. Drug Des Devel Ther. 2015 Sep 29;9:5421-31.
74. Pallàs M, Verdaguer E, Jordà EG, Jiménez A, Canudas AM, Camins A. Flavopiridol: an antitumor drug with potential application in the treatment of neurodegenerative diseases.

534 Med Hypotheses. 2005;64(1):120-3.
535 75. Gsell W, Conrad R, Hickethier M, Sofic E, Frölich L, Wichart I, Jellinger K, Moll G,
536 Ransmayr G, Beckmann H, et al. Decreased catalase activity but unchanged superoxide
537 dismutase activity in brains of patients with dementia of Alzheimer type. J Neurochem.
538 1995 Mar;64(3):1216-23.
539 76. Lovell MA, Ehmann WD, Butler SM, Markesbery WR. Elevated thiobarbituric acid-
540 reactive substances and antioxidant enzyme activity in the brain in Alzheimer's disease.
541 Neurology. 1995 Aug;45(8):1594-601.
542 77. Kim TH, Hong JM, Oh B, Cho YS, Lee JY, Kim HL, Shin ES, Lee JE, Park EK, Kim SY.
543 Genetic association study of polymorphisms in the catalase gene with the risk of
544 osteonecrosis of the femoral head in the Korean population. Osteoarthritis Cartilage. 2008
545 Sep;16(9):1060-6.
546 78. Mizwicki MT, Liu G, Fiala M, Magpantay L, Sayre J, Siani A, Mahanian M, Weitzman R,
547 Hayden EY, Rosenthal MJ, et al. 1α,25-dihydroxyvitamin D3 and resolvin D1 retune the
548 balance between amyloid-β phagocytosis and inflammation in Alzheimer's disease patients.
549 J Alzheimers Dis. 2013; 34(1):155-70.
550 79. Cifuentes RA, Murillo-Rojas J. Alzheimer's disease and HLA-A2: linking
551 neurodegenerative to immune processes through an in silico approach. Biomed Res Int.
552 2014;2014:791238.
553 80. Yang J, Li S, He XB, Cheng C, Le W. Induced pluripotent stem cells in Alzheimer's
554 disease: applications for disease modeling and cell-replacement therapy. Mol Neurodegener.
555 2016 May 17;11(1):39.
556 81. Xinzhong Li, Jintao Long, Taigang He, Robert Belshaw, and James Scott. Integrated
557 genomic approaches identify major pathways and upstream regulators in late onset
558 Alzheimer's disease. Sci Rep. 2015; 5: 12393.
559 82. Engidawork E, Gulesserian T, Yoo BC, Cairns N, Lubec G. Alteration of caspases and
560 apoptosis-related proteins in brains of patients with Alzheimer's disease. Biochem Biophys
561 Res Commun. 2001 Feb 16;281(1):84-93.
562 83. Reed JC. Mechanisms of apoptosis. Am J Pathol. 2000 Nov;157(5):1415-30.
563 84. Gopalraj RK, Zhu H, Kelly JF, Mendiondo M, Pulliam JF, Bennett DA, Estus S. Genetic
564 association of low density lipoprotein receptor and Alzheimer's disease. Neurobiol Aging.
565 2005; 26(1): 1-7.
566 85. Bowen RL, Isley JP, Atkinson RL. An association of elevated serum gonadotropin
567 concentrations and Alzheimer disease? J Neuroendocrinol. 2000 Apr;12(4):351-4.
568 86. Vingtdeux V, Davies P, Dickson DW, Marambaud P. AMPK is abnormally activated in
569 tangle- and pre-tangle-bearing neurons in Alzheimer's disease and other tauopathies. Acta

Neuropathol. 2011 Mar;121(3):337-49.
87. Greco SJ, Sarkar S, Johnston JM, Tezapsidis N. Leptin regulates tau phosphorylation and amyloid through AMPK in neuronal cells. Biochem Biophys Res Commun. 2009 Feb 27;380(1):98-104.
88. Salminen A, Kaarniranta K, Haapasalo A, Soininen H, Hiltunen M. AMP-activated protein kinase: a potential player in Alzheimer's disease. J Neurochem. 2011 Aug;118(4):460-74.
89. Semagacestat | Alzforum. Retrieved April 8, 2019, from https://www.alzforum.org/therapeutics/semagacestat.
90. Verubecestat | Alzforum. Retrieved April 8, 2019, from https://www.alzforum.org/therapeutics/verubecestat.
91. Atabecestat | Alzforum. Retrieved April 8, 2019, from https://www.alzforum.org/therapeutics/atabecestat.
92. Chang Y, Paramasivam M, Girgenti MJ, Walikonis RS, Bianchi E, LoTurco JJ. RanBPM regulates the progression of neuronal precursors through M-phase at the surface of the neocortical ventricular zone. Dev Neurobiol. 2010 Jan;70(1):1-15.
93. Tufail Y, Cook D, Fourgeaud L, Powers CJ, Merten K, Clark CL, Hoffman E, Ngo A, Sekiguchi KJ, O'Shea CC, Lemke G, Nimmerjahn A. Phosphatidylserine Exposure Controls Viral Innate Immune Responses by Microglia. Neuron. 2017 Feb 8;93(3):574-586.e8.
94. Rojiani MV, Alidina J, Esposito N, Rojiani AM. Expression of MMP-2 correlates with increased angiogenesis in CNS metastasis of lung carcinoma. Int J Clin Exp Pathol. 2010 Oct 16;3(8):775-81.

Table 1(on next page)

Main sources for data integration

1
2

Database	Category	Website	Reference
BioGRID	Protein-protein interaction database	https://thebiogrid.org/	[16]
HMDD	Disease-associated miRNAs, miRNA-associated targeted genes	http://www.cuilab.cn/hmdd	[17]
TAG	Tumor-associated genes	http://www.binfo.ncku.edu.tw/TAG	[18]
TarBase	miRNA-associated targeted genes	http://www.microrna.gr/tarbase	[19]
KEGG	Biological pathways	http://www.genome.jp/kegg	[20]
GO	Gene ontology, gene annotation and product	http://geneontology.org/	[21]
GeneBank	Gene expression data	https://www.ncbi.nlm.nih.gov/genbank	[22]

3
4

Table 2(on next page)

Data contained in current version miRDRN

1

	Disease	miRNA	miRNA regulated gene	RSP
Number of	119	207	389	6,973,875

2
3

Table 3(on next page)

Result for sample Single Search: disease, colorectal neoplasm; tissue type, colorectal

1

	Disease
Disease Name	colorectal neoplasms
Tissue Filter	colorectal tumor
Associated miRNAs (total:33)	hsa-mir-491, hsa-mir-185, hsa-mir-20a, hsa-mir-221, hsa-mir-199a, hsa-mir-34a, hsa-mir-199b, hsa-mir-34c, hsa-mir-34b, hsa-mir-148a, hsa-mir-342, hsa-mir-21, hsa-mir-499a, hsa-let-7c, hsa-mir-148b, hsa-mir-1915, hsa-mir-17, hsa-mir-320a, hsa-mir-200c, hsa-mir-143, hsa-mir-139, hsa-mir-103a, hsa-mir-103b, hsa-mir-107, hsa-mir-497, hsa-mir-106a, hsa-mir-429, hsa-mir-7, hsa-mir-362, hsa-mir-330, hsa-mir-367, hsa-mir-339, hsa-mir-133a
Targeted genes (total:37)	BCL2L1, RHOA, CDC42, BNIP2, CDKN1C, AXL, MYC, BCL2, DNMT1, RHOB, FOXO4, PDCD4, MMP11, PBX3, CCKBR, CCL20, RND3, NRP1, ZEB1, CTNNB1, MACC1, IGF1R, DAPK1, KLF4, RAP1B, TGFBR2, SOX2, YY1, RBL2, E2F1, USF2, PTPN1, RYR3, PLRG1, RFFL, DNMT3A, KRAS
No. of RSPs	45565
No. of distinct genes	3079

2

3

Table 4(on next page)

Statistics and gene information in the Network-1, the largest connected sub-network of the CRC-associated miRNA regulatory network

1

		Number	Item Set
Network-1	miRNAs	6	hsa-mir-199a, hsa-mir-34a, hsa-mir-199b, hsa-mir-139, hsa-mir-497, hsa-mir-106a
	Target genes	4	AXL, IGF1R, RAP1B, TGFBR2
	Gene Set (including target genes)	56	AXL, CSK, TNK2, LCK, PRKACA, FGR, MAPK15, IGF1R, MERTK, ERBB2, PTK2, EGFR, JAK2, JAK1, PRKCD, TEC, EPHB2, PHKG2, ROR1, FES, MAP3K12, RAP1B, MST4, PAK1, LRRK1, MAP2K3, CDK11B, ACVR1, TGFBR2, RIOK2, TGFBR1, MAP3K7, NEK8, NUAK2, OXSR1, CDK1, ACVRL1, MKNK2, STK35, CDK17, EIF2AK4, DAPK2, EIF2AK1, TSSK4, ZAK, MAP2K6, SIK3, VRK2, PINK1, TAOK2, TNIK, MAPK6, PRKACB, WNK1, PAK6, PKMYT1

2
3

Table 5(on next page)

Known, literature supported, and potential novel CRC-related genes

1

		Number	Item Set
Gene Set（Network-1）	Known CRC genes	22	AXL, LCK, FGR, IGF1R, MERTK, ERBB2, PTK2, EGFR, JAK2, JAK1, EPHB2, FES, PAK1, MAP2K3, ACVR1, TGFBR2, TGFBR1, CDK1, EIF2AK4, DAPK2, MAP2K6, PAK6
	Reference supported [25-52]	26	CSK, TNK2*, MAPK15, PRKCD, TEC, PHKG2, ROR1, RAP1B, MST4, CDK11B, MAP3K7, NEK8, NUAK2, ACVRL1, MKNK2, STK35, ZAK, SIK3, VRK2, PINK1, TAOK2, TNIK*, MAPK6, PRKACB, PKMYT1, WNK1
	Potential novel cancer-related gene	8	PRKACA, MAP3K12, LRRK1, RIOK2, OXSR1, CDK17, EIF2AK1, TSSK4

2 *Known target genes for the treatment in CRC

3

4

Table 6(on next page)

Results for the AD-T2D comorbidity study

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27699v1 | CC BY 4.0 Open Access | rec: 2 May 2019, publ: 2 May 2019

1

	Disease 1	Disease 2	Comorbidity
Disease Name	AD	T2D	AD/T2D
Tissue Filter	brain	pancreas	brain/pancreas
Associated-miRNA	hsa-mir-29a, hsa-mir-195, hsa-mir-146a	hsa-mir-144	hsa-mir-29a, hsa-mir-195, hsa-mir-146a, hsa-mir-144
Targeted gene	NAV3, BACE1, CFH	IRS1	NAV3, BACE1, CFH, IRS1
Regulatory sub-pathways	644	3908	4552
Total no. of genes	633	2187	2320
No. of common genes	-	-	500

2
3

Table 7 (on next page)

Known, literature supported, and potential novel AD-T2D comorbid genes

The 210 known AD target genes were built by integrating gene lists from AlzGene (http://www.alzforum.org/genetics) [54], AlzBIG (http://alz.big.ac.cn/) [55] and AlzBase (http://alz.big.ac.cn/alzBase/home) [56]; the 497 known T2D targets contains 497 genes were from T-HOD (http://bws.iis.sinica.edu.tw/THOD/) [57]. *Known AD and T2D target; #No literature support.

1

		Number of targets in comorbidity gene set (500)	Comorbid Genes (references in square brackets)
Known data	AD target (210)	8	ALOX5 [58-60], APP*, BIN1 [61], CHGB [62], VWF*, NEFL [63,64], LETMD1#, CELF1 [65,66]
	T2D target (497)	14	TCF7L2 [67-70], APOA1 [71-73], VWF*, CDKN2B [74], CAT [75-77], ITGB2 [78,79], ISL1 [80], POLD3#, APP*, NFKBIB [81], GNA12#, DEDD [82,83], LDLR [84,85], PRKAB1 [86-88]

2

Figure 1

A regulatory sub-pathway (RSP)

Given a disease-associated miRNA, M, and its target gene T, the linked sequence (M, T, G_1, G_2) is an RSP associated with M, where G_1 is a protein interacting (according to PPI data) with T, and G_2 is a protein interacting with G_1. In the text, G_1 (G_2) is said to have a level 1 (level 2) PPI with T.

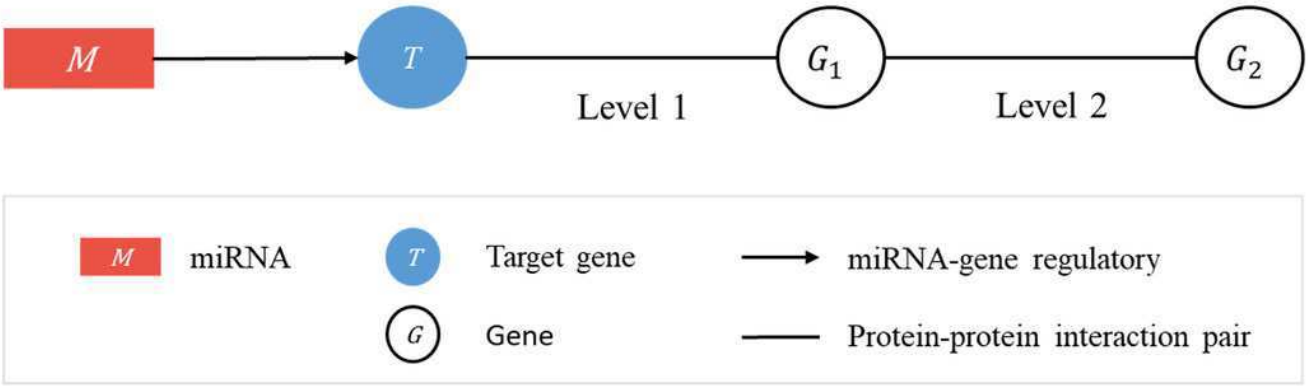


Figure 2

Schematic construction of disease specific miRNA regulatory network (RRN)

For a given disease there may be more than one miRNA associated with it, and each disease-associated miRNA may have one or more target genes. After all the RSPs having the from (M, T, G_1, G_2) are constructed, an RRN is built from entire set of constructed RSPs by linking (pairs of) genes/proteins from different RSPs whenever they interact according to PPI data.

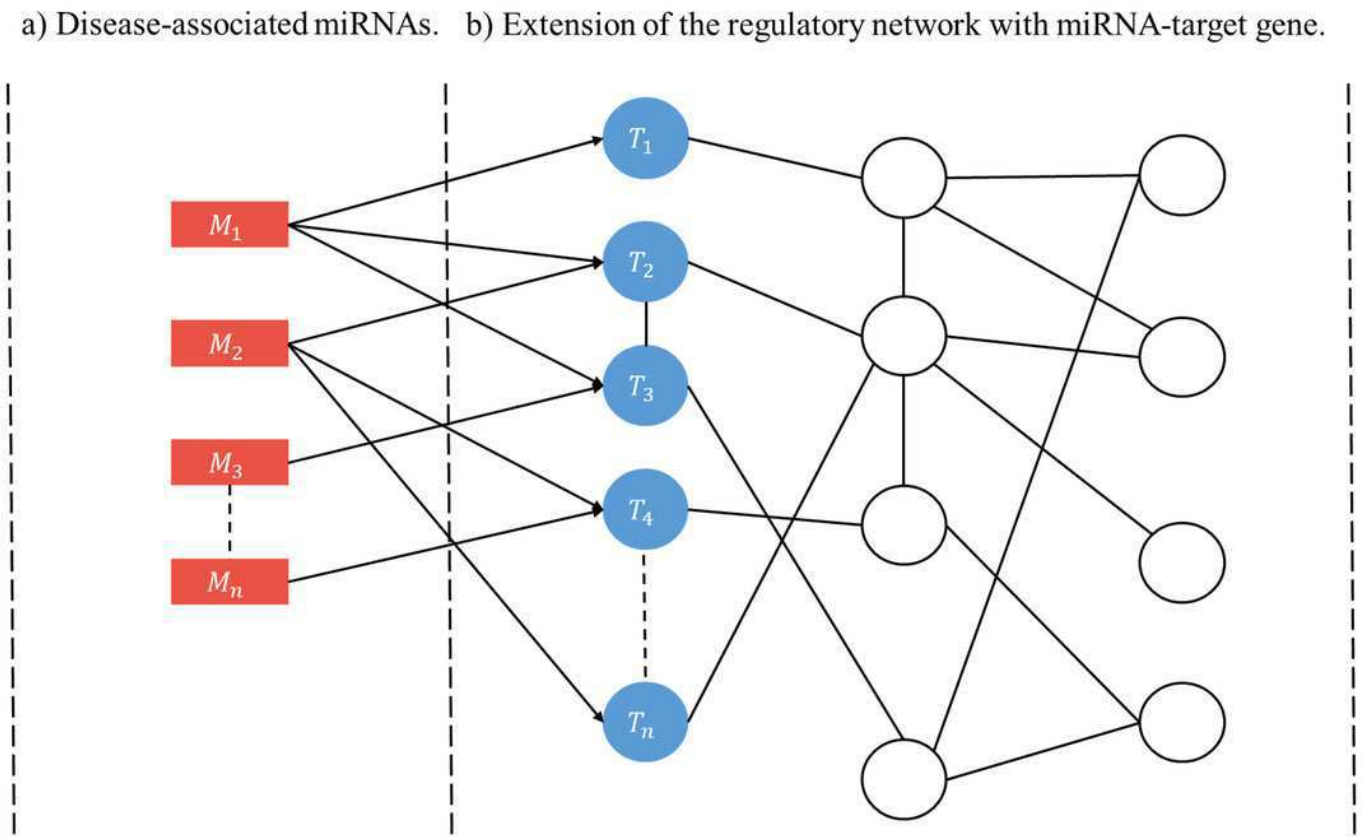


Figure 3

Entrance page of miRDRN

User may select “Single Search” to explore a single disease or “Comorbidity Search” to explore a disease-pair.

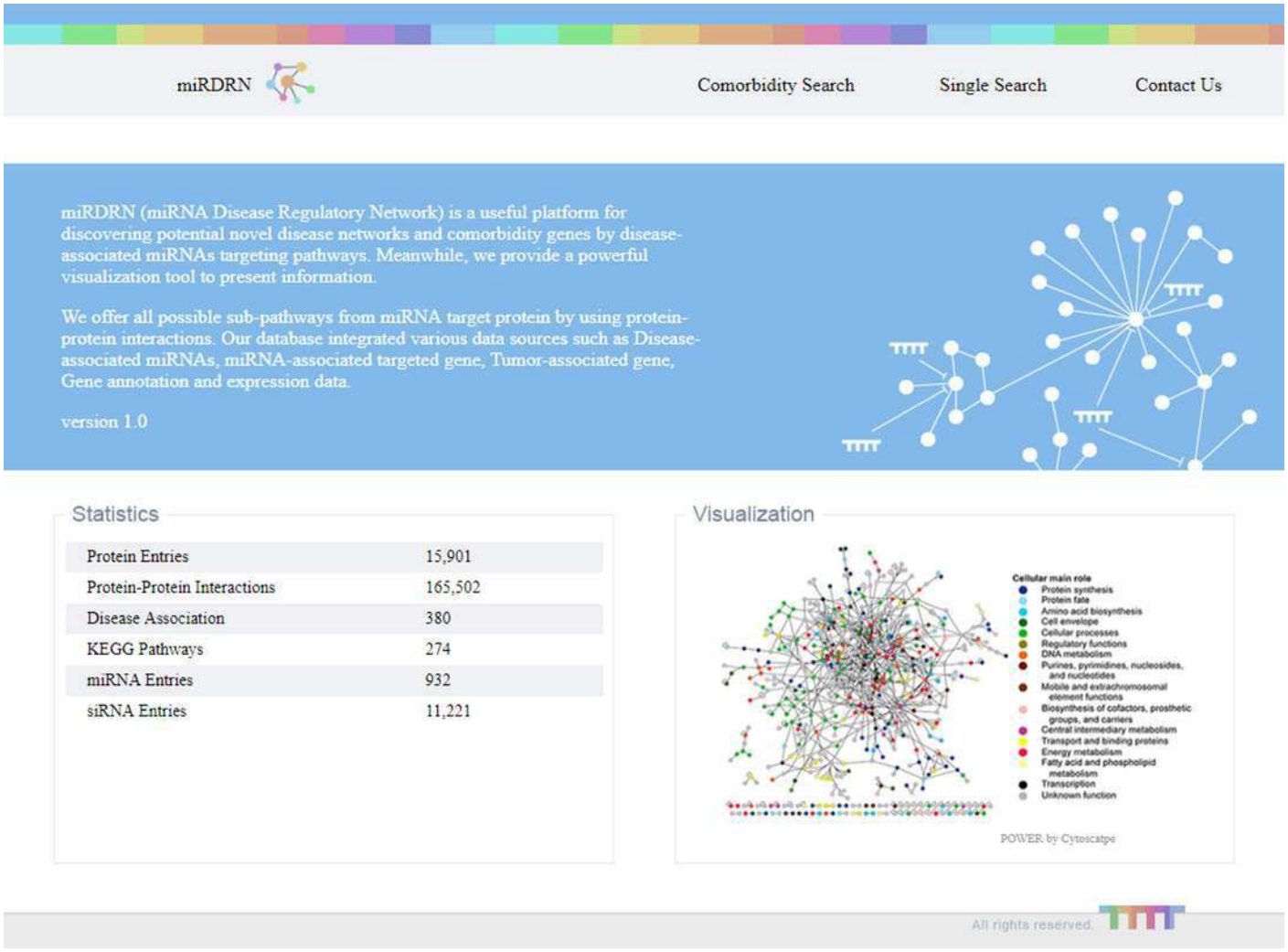


Figure 4

Query interface of Comorbidity Search

User is required to select two diseases, Disease 1 (“Alzheimer’s Disease” selected) and Disease 2 (“Type 2” selected), their respective tissue/tumor types (“bran” and “pancreas”, respectively), pathway ranking method (by biological processes (BP) or molecular functions (MF)), and *p*-value threshold (0.005 for both diseases). There are also four optional filters regarding gene property (none selected).

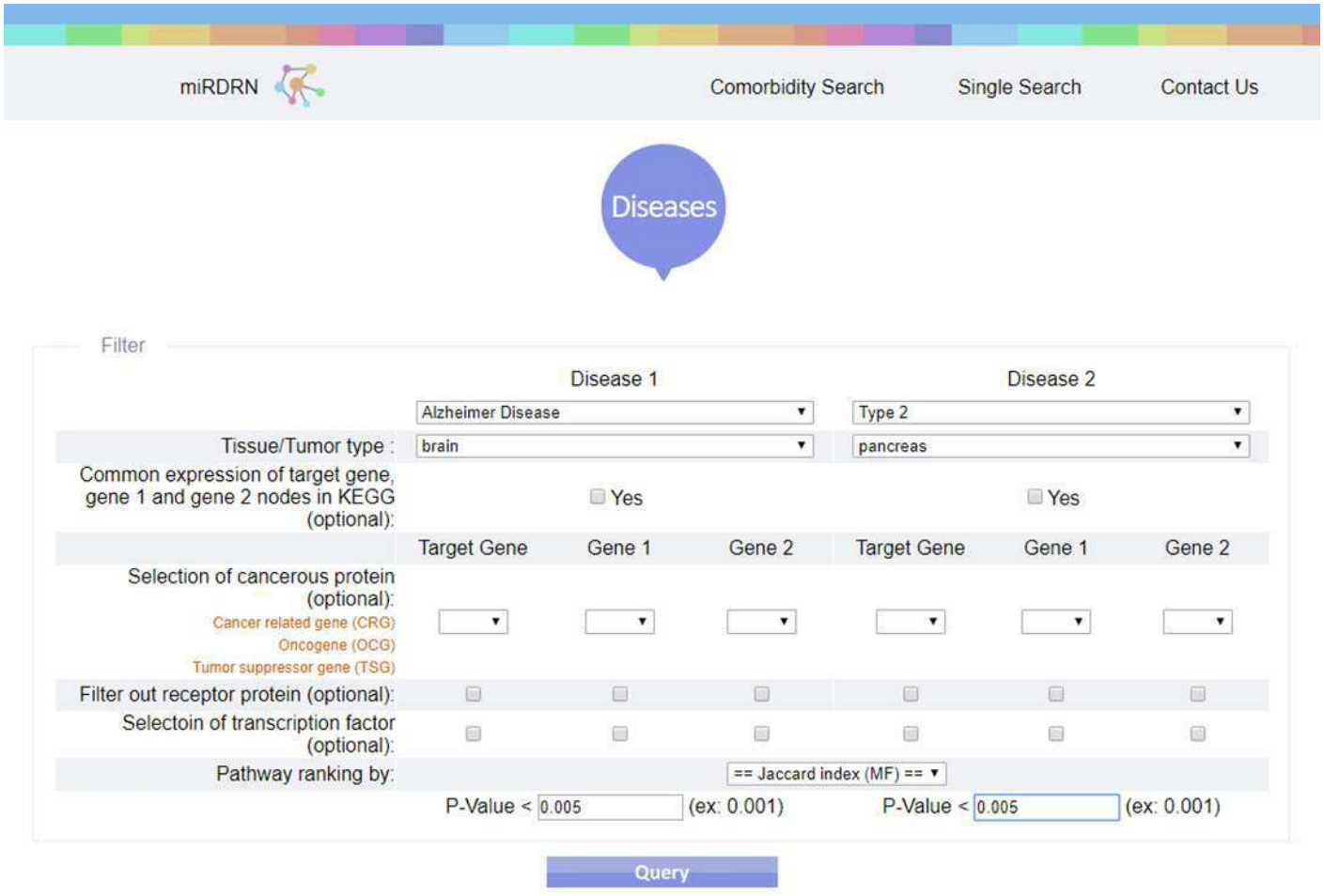


Figure 5

Result for miRNAs in Alzheimer's Disease-Type 2 Diabetes Comorbidity Search

Search result for miRNAs associated with Disease 1 (Alzheimer's Disease in this example) and Disease 2 (Type 2 Diabetes) and literature source (blue area) and list of gene(s) targeted by each miRNA (green area). For each gene the gene symbol and its OMIM id are given, as well as information on whether the protein it encodes has a cancerous protein tag: CRG, cancer related gene; OCG, oncogene; TSG, tumor suppressor gene.

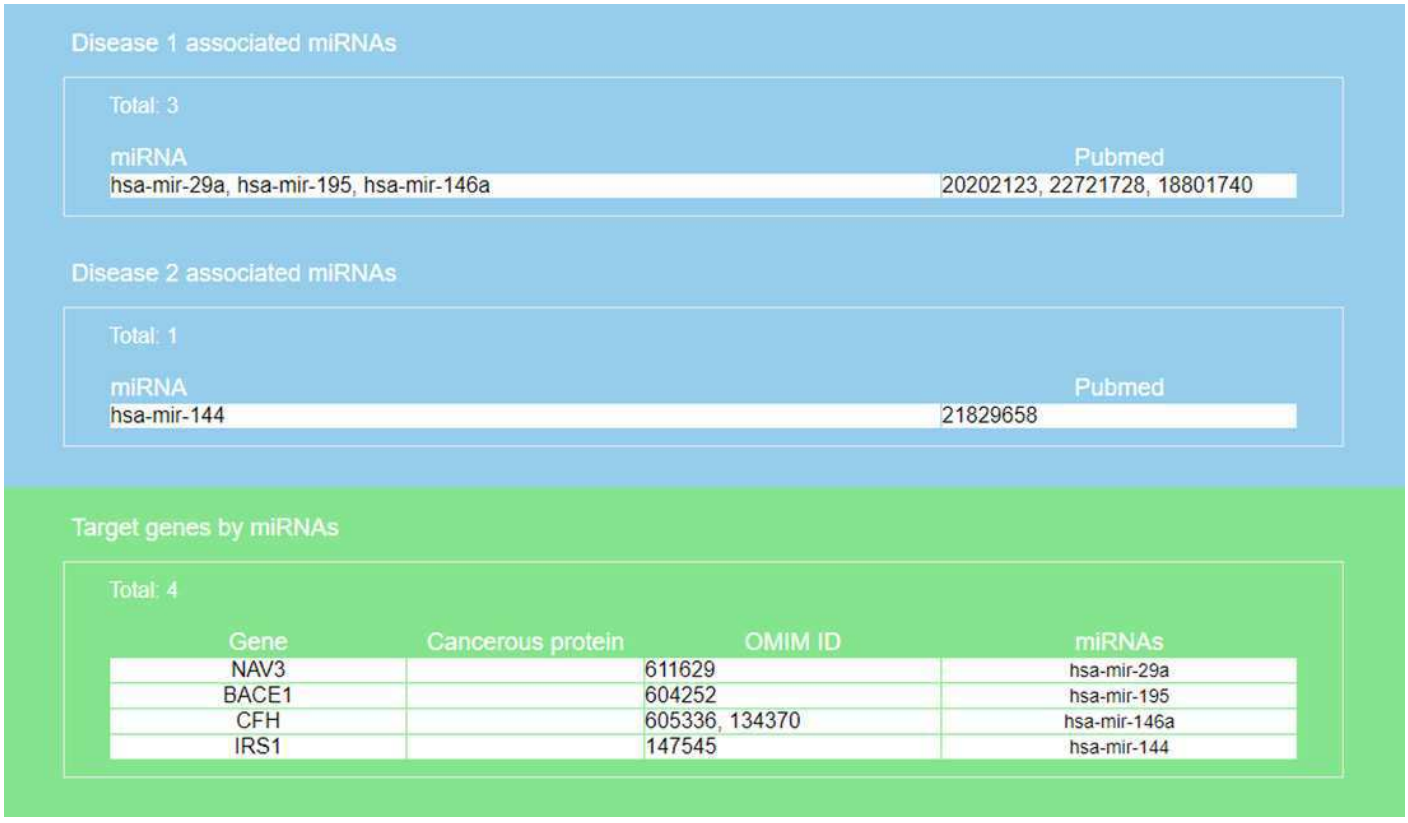


Figure 6

Result on RSP in Alzheimer's Disease-Type 2 Diabetes Comorbidity Search

RSPs are listed in descending order (column 1) by *p*-value (column 6). Columns 2-4 give the symbols of genes in the sequence (T, G_1, G_2). Column 5 gives known pathways, such as a KEGG pathway, of which (T, G_1, G_2) is a part. On first appearance, all RSPs (4552 in this example) are listed on multiple pages. There are two options for displaying/using a smaller RSP set: “Gene filter”, where user can restrict the set to only those RSPs containing a specified gene, and “Show top … sub-pathways”, where user can ask for only the N-top RSPs having the smallest *p*-values be listed and used for network construction.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27699v1 | CC BY 4.0 Open Access | rec: 2 May 2019, publ: 2 May 2019

miRNA regulated sub pathways

All pathways | Export Data

Total: 4552

List view: Gene filter (Please input a gene symbol, ex:BCL2),

Sort by P-Value , Show top 4552 sub-pathways GO (Number of list:4552)

| Next >>

No.	Target Gene	Gene 1	Gene 2	Common Pathway	P-Value
1	BACE1 RC	PSEN1 RC	PSMA5		9.443511e-14
2	BACE1 RC	PSEN1 RC	PSMB1		9.443511e-14
3	BACE1 RC	PSEN1 RC	STAMBPL1		2.761702e-10
4	BACE1 RC	PSEN1 RC	CASP1		5.522146e-10
5	IRS1	YWHAG	YWHAH	hsa04151	7.887880e-10
6	BACE1 RC	PSEN1 RC	CTNNA1		9.661552e-10
7	IRS1	CBLB CRG	ASAP2 RC		3.310267e-9
8	IRS1	CBLB CRG	NR2C2 TF		3.310267e-9
9	IRS1	PIK3R1	YWHAG	hsa04151	8.731441e-9
10	IRS1	PTK2	TRIO OCG		1.125868e-8
11	IRS1	PTK2	RET OCG RC		1.125868e-8
12	IRS1	PTK2	MAPK8IP3		1.125868e-8
13	IRS1	PTK2	EPHB2 TSG RC		1.125868e-8
14	IRS1	PTK2	PHKG2		1.125868e-8
15	IRS1	AKT1 OCG	PFKFB2		1.790201e-8
16	IRS1	AKT1 OCG	WNK1		1.790201e-8
17	IRS1	AKT1 OCG	PDK2		1.790201e-8
18	IRS1	AKT1 OCG	CLK2		1.790201e-8
19	IRS1	AKT1 OCG	PINK1		1.790201e-8
20	IRS1	AKT1 OCG	CKB		1.790201e-8
21	IRS1	AKT1 OCG	PI4K2B RC		1.790201e-8
22	IRS1	AKT1 OCG	MARK2 RC		1.790201e-8
23	IRS1	AKT1 OCG	PAK6		1.790201e-8
24	IRS1	JAK2 RC	TEC OCG RC		3.808482e-8
25	IRS1	JAK2 RC	FES OCG		3.808482e-8
26	IRS1	INSR RC	MOK		3.808482e-8
27	CFH	ITGAM RC	ITGB2 RC	hsa05150	6.060787e-8
28	IRS1	PIK3R1	PIK3C2B RC		9.565237e-8
29	IRS1	PTK2	BBS10		2.251149e-7
30	IRS1	CBLB CRG	TANK		3.116147e-7
31	IRS1	SOCS1	SOCS2	hsa04930, hsa04910	4.327160e-7
32	IRS1	PIK3R1	FYB		6.060787e-7
33	IRS1	YWHAG	PARD3		6.489754e-7
34	IRS1	YWHAG	CDC5L OCG TF		6.489754e-7
35	IRS1	GRB10	DOK1		7.787705e-7
36	BACE1 RC	MMP2 RC	MMP25 RC		9.084275e-7
37	BACE1 RC	MMP2 RC	MMP17 RC		9.084275e-7
38	IRS1	YWHAG	ARHGEF16		9.084275e-7
39	IRS1	IRS2	IL4R RC		9.084275e-7
40	IRS1	AKT1 OCG	IBTK RC		9.086691e-7

Figure 7

Result on comorbidity genes in Alzheimer's Disease-Type 2 Diabetes Comorbidity Search

Genes common to some RSPs of both diseases are listed, together with information on cancer genes status, OMIM Id, and KEGG pathway.

Comorbidity Gene

Export Data

Total: 500 | Next >>

No.	Gene	Cancerous protein	OMIM ID	KEGG
1	A2M		103950	hsa04610
2	ABCF1		603429	
3	ACHE		100740	hsa00564, hsa04725
4	ACO1		100880	hsa00020, hsa00630, hsa00300
5	AFG3L1P		603020	
6	ALOX5		152390	hsa04726, hsa05145, hsa00590, hsa04913
7	BIN1	TSG	601248	
8	ANXA11		602572	
9	APBB1		602709	hsa05010
10	APBB2		602710	
11	APLP1		104775	
12	APOA1		107680	hsa03320, hsa04977, hsa04975, hsa05143
13	APP		605378, 100070, 104760	hsa04726, hsa05010
14	ARL4D		600732	
15	ASS1		603470	hsa00250, hsa00330
16	RERE		605226	
17	ATP5A1		164360	hsa05012, hsa05016, hsa00190, hsa05010
18	ALDH7A1		107323	hsa00260, hsa00380, hsa00330, hsa00280, hsa00410, hsa00300, hsa00561, hsa00310, hsa00340, hsa00640, hsa00053, hsa00620, hsa00071, hsa00010
19	ATP6V1E1		108746	hsa05110, hsa04966, hsa05323, hsa04721, hsa05120, hsa00190, hsa04145
20	BLMH		602403	
21	BNIP1		603291, 612478	hsa04130
22	BTF3		613595	
23	MRPL49		606866	
24	CARS		123859	hsa00970
25	CAT		115500, 607424	hsa05014, hsa00380, hsa00630, hsa04146

Figure 8

Display of a sub-RRN built from a subset of RSPs determined by the user using options available in the interface shown in Figure 6

The option "Show top-70" RSPs (by p-value) was used. When the mouse is placed on a node in the displayed RRN, a small text window opens showing the name of the node/gene and annotations from GO, OMIM, KEGG, and GeneBank databases.

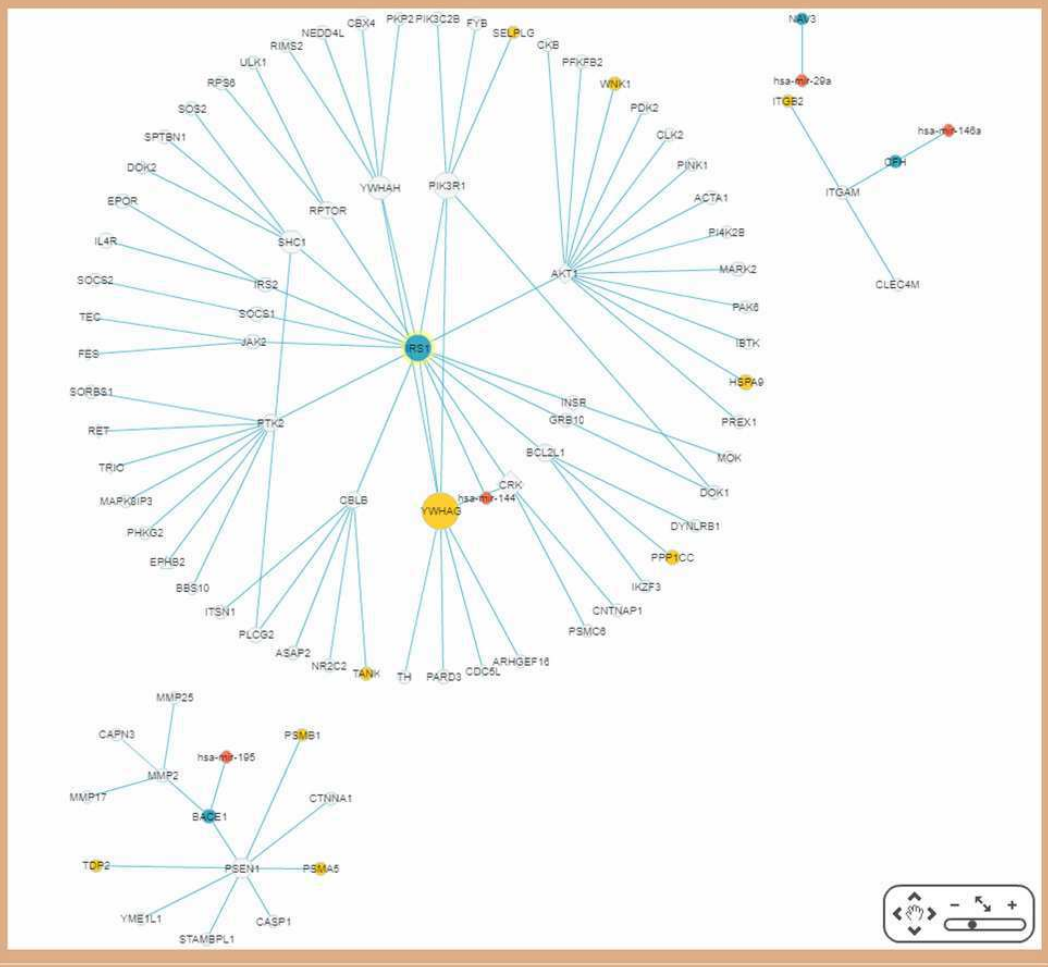
Disease specific miRNAs regulated network

Cancerous gene : ◇ / OCG △ / TSG ▭ / BOTH

Color labeling : ● / miRNA ● / Target Gene ● / Comorbidity Gene

Layout : Radial

(If you re-setting the sub-pathways list view, please select this layout again)



Information for selected node:

Gene:	IRS1
Cancerous protein:	
OMIM ID:	147545
KEGG pathway:	hsa04722 Neurotrophin signaling pathway - Homo sapiens (human)
	hsa04151 PI3K-Akt signaling pathway - Homo sapiens (human)
	hsa04930 Type II diabetes mellitus - Homo sapiens (human)
	hsa04150 mTOR signaling pathway - Homo sapiens (human)
	hsa04920 Adipocytokine signaling pathway - Homo sapiens (human)

Cellular component	Biological process	Molecular function
GO:0005901	caveola	
GO:0005829	cytosol	
GO:0005899	insulin receptor complex	
GO:0005634	nucleus	
GO:0005737	cytoplasm	
GO:0005886	plasma membrane	

Figure 9

A partial miRNA regulatory network (RRN) for colorectal neoplasm

The RRN is constructed from the top 70 RSPs by *p*-value for colorectal neoplasm, tissue type, colorectal tumor. A link indicates a miRNA-target relation or a PPI; red circle, miRNA; blue circle, miRNA target gene; yellow circle, non-target gene; diamond, oncogene; triangle, tumor suppressor gene.

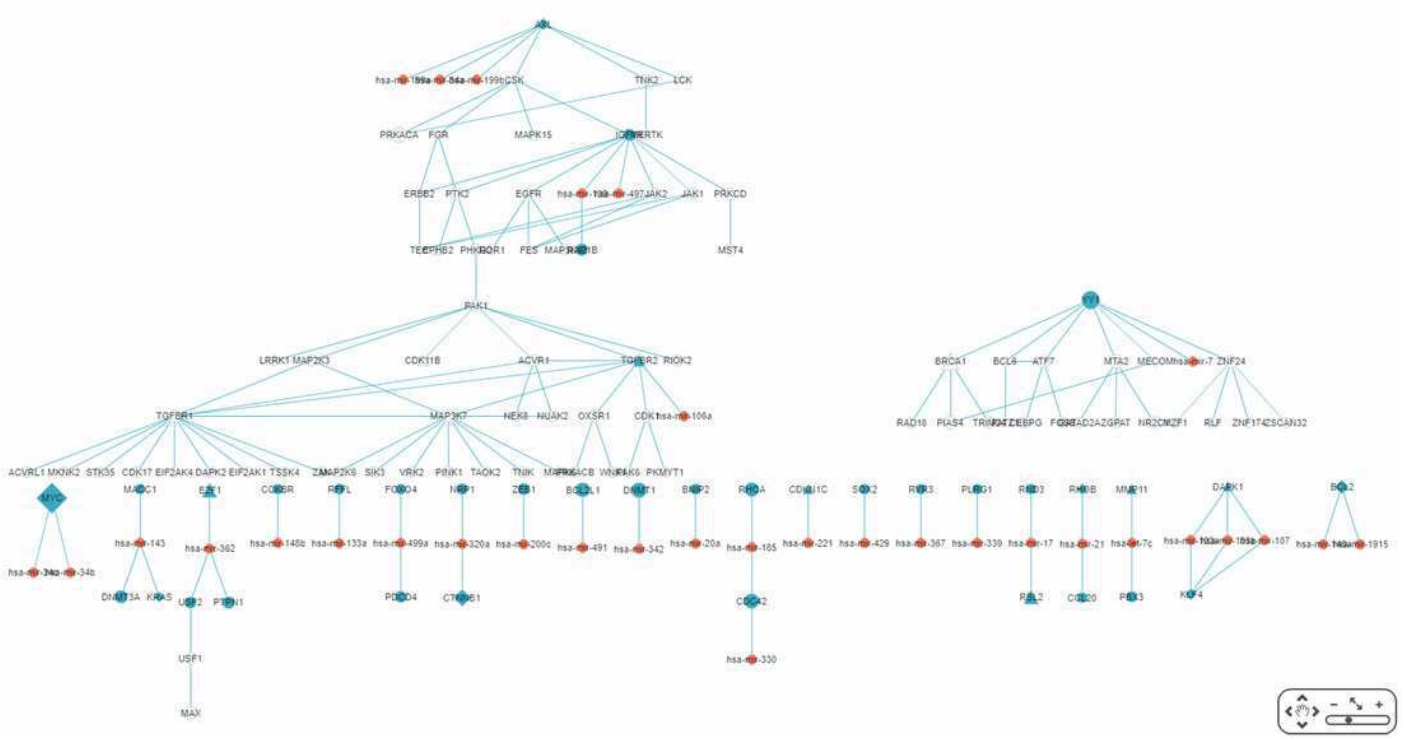


Figure 10

The sub-RRN Network-1

This largest connected sub-RRN for colorectal neoplasm (constructed from the top 70 RSPs by *p*-value), containing six miRNAs targeting four genes connected to 52 other genes, is itself composed of two parts, one 28 nodes (five miRNAs targeting three genes) and the other 34 nodes (one miRNA targeting one gene), connected by a single link.

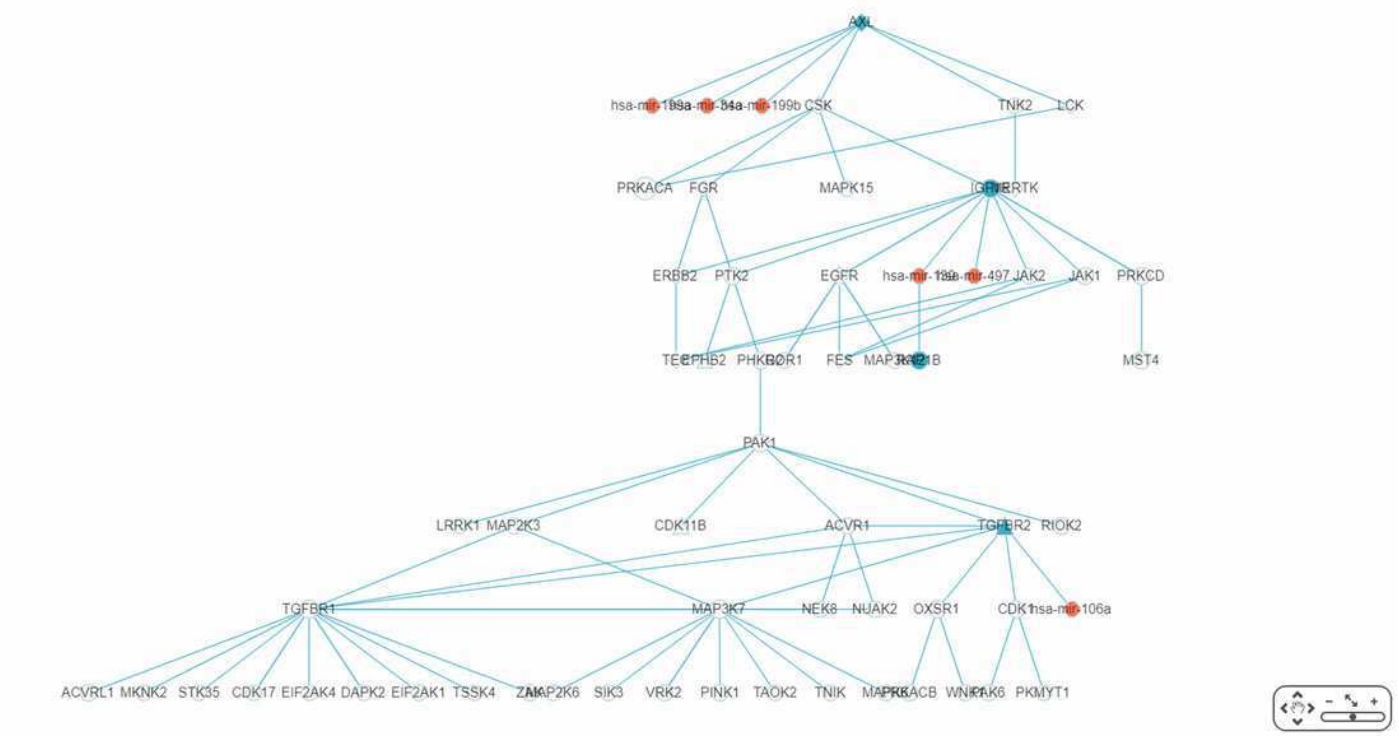
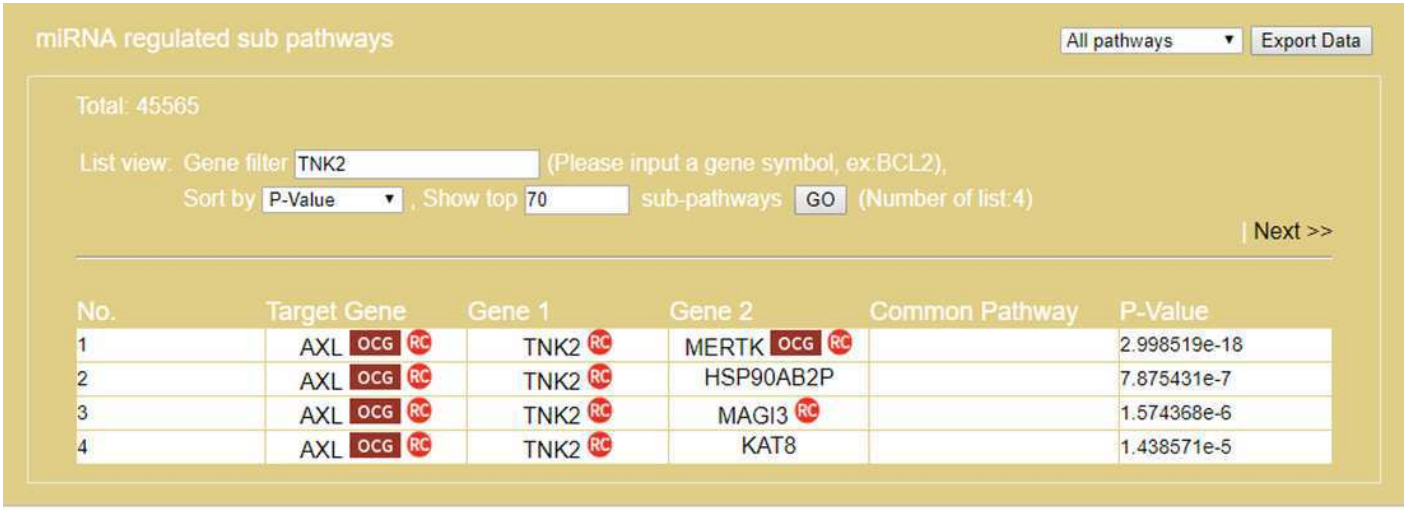


Figure 11

A sub-RRN of CRC obtained by using *TNK2* as a gene filter

The RRN contains the target gene *AXL* regulated by three miRNAs, hsa-mir-199b, hsa-mir-34a, hsa-mir-199a, and linked by PPI to *TNK2*, itself linked by PPI to four other genes *AXL*(OCG), *MAGI3*, *HSP90AB2P*, *MERTK*(OCG), *KAT8*.



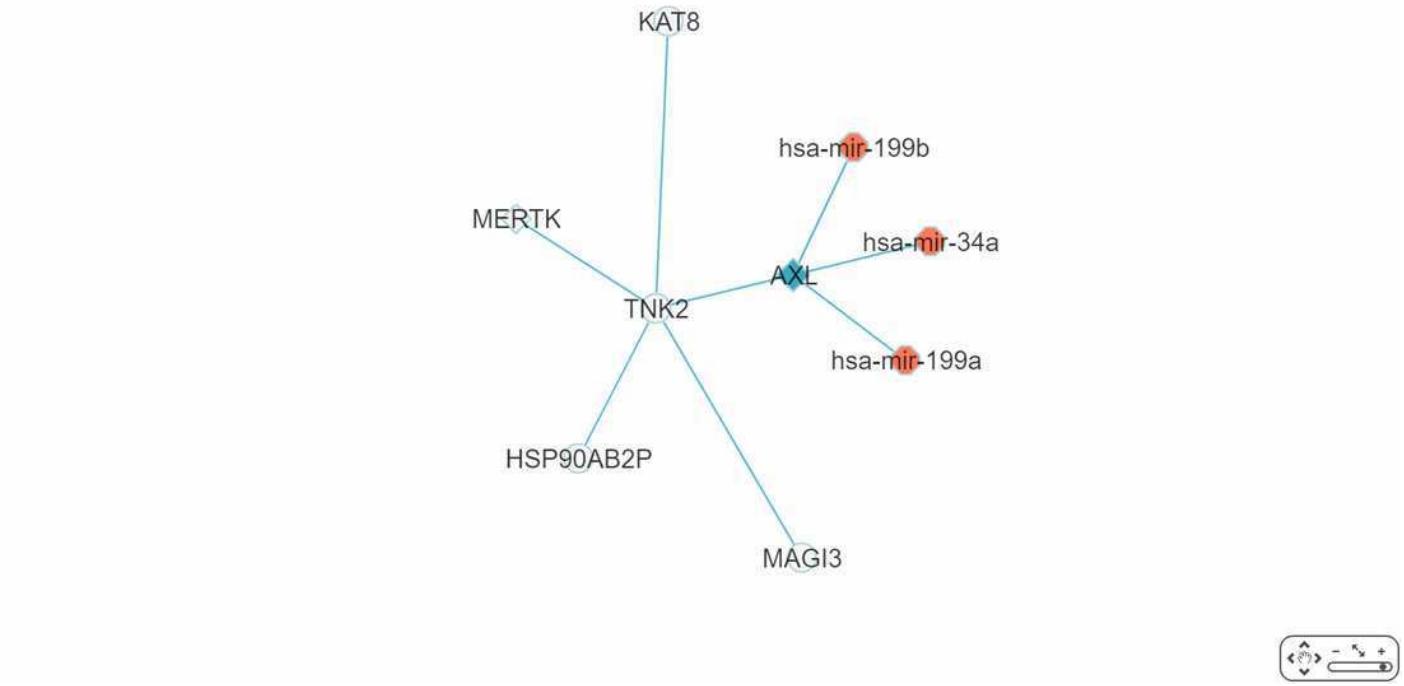


Figure 12

A miRNA regulatory sub-network centered on the AD-associated gene *BACE1*

The genes *PSEN1*, *NCSTN*, *RANBP9*, *PLSCR1*, *MMP2*, and *FURIN* are shown to be immediately downstream to, i.e., have level 1 PPI with, *BACE1*.